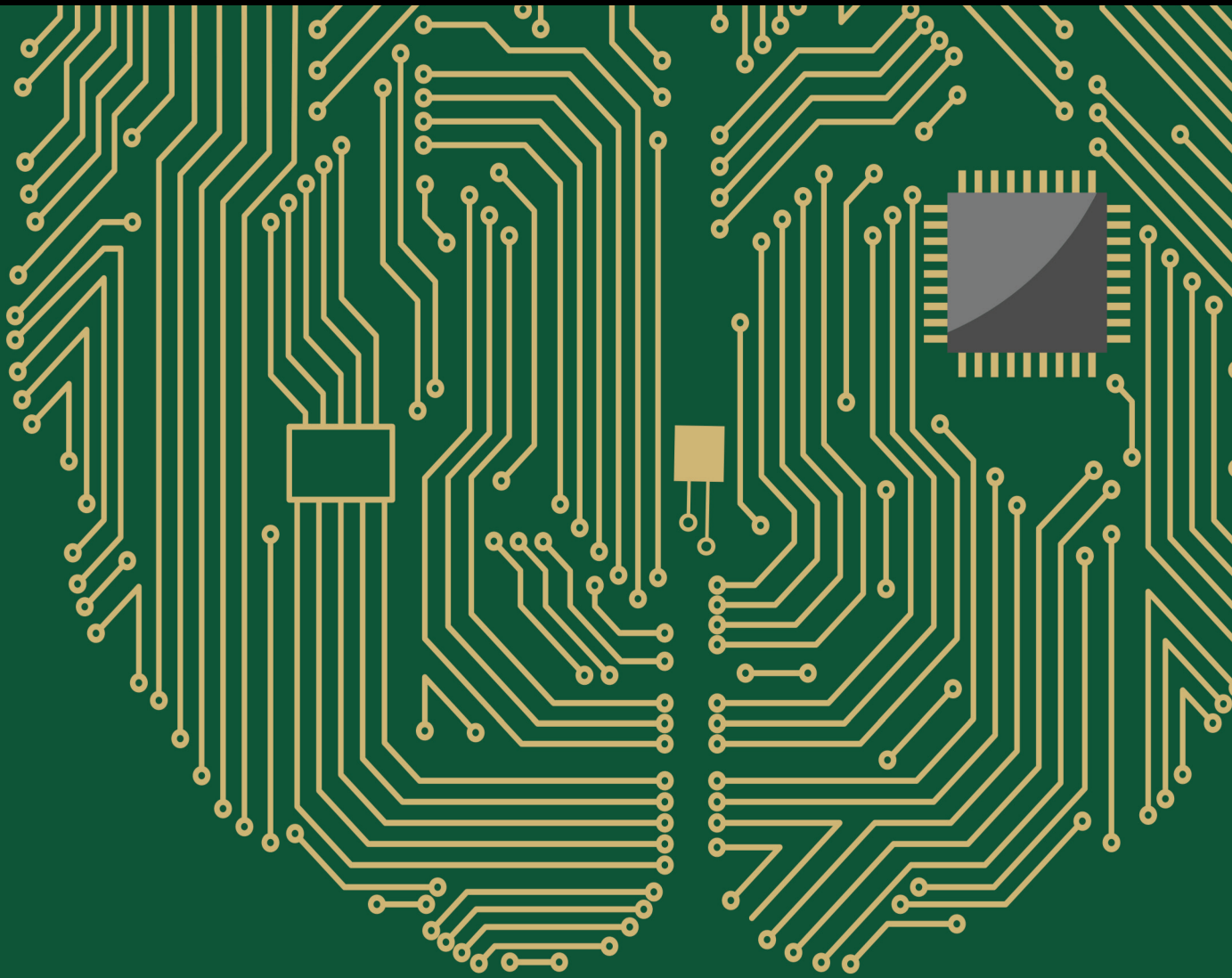


Computational Intelligence for Intelligent Human-Machine Systems

Lead Guest Editor: Zhongxu Hu

Guest Editors: Jie Liu, Chen Lv, and Yang Xing





Computational Intelligence for Intelligent Human-Machine Systems

Computational Intelligence and Neuroscience

**Computational Intelligence for
Intelligent Human-Machine Systems**

Lead Guest Editor: Zhongxu Hu

Guest Editors: Jie Liu, Chen Lv, and Yang Xing



Copyright © 2023 Hindawi Limited. All rights reserved.

This is a special issue published in "Computational Intelligence and Neuroscience." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chief Editor

Andrzej Cichocki, Poland

Associate Editors

Arnaud Delorme, France
Cheng-Jian Lin , Taiwan
Saeid Sanei, United Kingdom

Academic Editors









Mohamed Abd Elaziz , Egypt
Tariq Ahanger , Saudi Arabia
Muhammad Ahmad, Pakistan
Ricardo Aler , Spain
Nouman Ali, Pakistan
Pietro Aricò , Italy
Lerina Aversano , Italy
Ümit Ağbulut , Turkey
Najib Ben Aoun , Saudi Arabia
Surbhi Bhatia , Saudi Arabia
Daniele Bibbo , Italy
Vince D. Calhoun , USA
Francesco Camastra, Italy
Zhicheng Cao, China
Hubert Cecotti , USA
Jyotir Moy Chatterjee , Nepal
Rupesh Chikara, USA
Marta Cimitile, Italy
Silvia Conforto , Italy
Paolo Crippa , Italy
Christian W. Dawson, United Kingdom
Carmen De Maio , Italy
Thomas DeMarse , USA
Maria Jose Del Jesus, Spain
Arnaud Delorme , France
Anastasios D. Doulamis, Greece
António Dourado , Portugal
Sheng Du , China
Said El Kafhali , Morocco
Mohammad Reza Feizi Derakhshi , Iran
Quanxi Feng, China
Zhong-kai Feng, China
Steven L. Fernandes, USA
Agostino Forestiero , Italy
Piotr Franaszczuk , USA
Thippa Reddy Gadekallu , India
Paolo Gastaldo , Italy
Samanwoy Ghosh-Dastidar, USA


Manuel Graña , Spain
Alberto Guillén , Spain
Gaurav Gupta, India
Rodolfo E. Haber , Spain
Usman Habib , Pakistan
Anandakumar Haldorai , India
José Alfredo Hernández-Pérez , Mexico
Luis Javier Herrera , Spain
Alexander Hošovský , Slovakia
Etienne Hugues, USA
Nadeem Iqbal , Pakistan
Sajad Jafari, Iran
Abdul Rehman Javed , Pakistan
Jing Jin , China
Li Jin, United Kingdom
Kanak Kalita, India
Ryotaro Kamimura , Japan
Pasi A. Karjalainen , Finland
Anitha Karthikeyan, Saint Vincent and the Grenadines
Elpida Keravnou , Cyprus
Asif Irshad Khan , Saudi Arabia
Muhammad Adnan Khan , Republic of Korea
Abbas Khosravi, Australia
Tai-hoon Kim, Republic of Korea
Li-Wei Ko , Taiwan
Raşit Köker , Turkey
Deepika Koundal , India
Sunil Kumar , India
Fabio La Foresta, Italy
Kuruva Lakshmana , India
Maciej Lawrynczuk , Poland
Jianli Liu , China
Giosuè Lo Bosco , Italy
Andrea Loddo , Italy
Kezhi Mao, Singapore
Paolo Massobrio , Italy
Gerard McKee, Nigeria
Mohit Mittal , France
Paulo Moura Oliveira , Portugal
Debajyoti Mukhopadhyay , India
Xin Ning , China
Nasimul Noman , Australia
Fivos Panetsos , Spain



Evgeniya Pankratova , Russia
Rocío Pérez de Prado , Spain
Francesco Pistolesi , Italy
Alessandro Sebastian Podda , Italy
David M Powers, Australia
Radu-Emil Precup, Romania
Lorenzo Putzu, Italy
S P Raja, India
Dr.Anand Singh Rajawat , India
Simone Ranaldi , Italy
Upaka Rathnayake, Sri Lanka
Navid Razmjooy, Iran
Carlo Ricciardi, Italy
Jatinderkumar R. Saini , India
Sandhya Samarasinghe , New Zealand
Friedhelm Schwenker, Germany
Mijanur Rahaman Seikh, India
Tapan Senapati , China
Mohammed Shuaib , Malaysia
Kamran Siddique , USA
Gaurav Singal, India
Akansha Singh , India
Chiranjibi Sitaula , Australia
Neelakandan Subramani, India
Le Sun, China
Rawia Tahrir , Iraq
Binhua Tang , China
Carlos M. Travieso-González , Spain
Vinh Truong Hoang , Vietnam
Fath U Min Ullah , Republic of Korea
Pablo Varona , Spain
Roberto A. Vazquez , Mexico
Mario Versaci, Italy
Gennaro Vessio , Italy
Ivan Volosyak , Germany
Leyi Wei , China
Jianghui Wen, China
Lingwei Xu , China
Cornelio Yáñez-Márquez, Mexico
Zaher Mundher Yaseen, Iraq
Yugen Yi , China
Qiangqiang Yuan , China
Miaolei Zhou , China
Michal Zochowski, USA
Rodolfo Zunino, Italy



Contents


Retracted: Human Resource Demand Prediction and Configuration Model Based on Grey Wolf Optimization and Recurrent Neural Network
Computational Intelligence and Neuroscience
Retraction (1 page), Article ID 9890474, Volume 2023 (2023)




AppraisalCloudPCT: A Computational Model of Emotions for Socially Interactive Robots for Autistic Rehabilitation
Ting Yan , Shengzhao Lin , Jinfeng Wang , Fuhao Deng , Zijian Jiang , Gong Chen ,
Jionglong Su , and Jiaming Zhang 
Research Article (25 pages), Article ID 5960764, Volume 2023 (2023)



Path Planning Algorithm for Unmanned Surface Vessel Based on Multiobjective Reinforcement Learning
Caimei Yang, Yingqi Zhao, Xuan Cai, Wei Wei, Xingxing Feng, and Kaibo Zhou 
Research Article (14 pages), Article ID 2146314, Volume 2023 (2023)

A Framework and Algorithm for Human-Robot Collaboration Based on Multimodal Reinforcement Learning
Zeyuan Cai, Zhiquan Feng , Liran Zhou, Changsheng Ai, Haiyan Shao, and Xiaohui Yang 
Research Article (13 pages), Article ID 2341898, Volume 2022 (2022)

Analysis on the Bus Arrival Time Prediction Model for Human-Centric Services Using Data Mining Techniques
N. Shanthi, Sathishkumar V E , K. Upendra Babu, P. Karthikeyan, Sukumar Rajendran, and Shaikh Muhammad Allayear 
Research Article (13 pages), Article ID 7094654, Volume 2022 (2022)

[Retracted] Human Resource Demand Prediction and Configuration Model Based on Grey Wolf Optimization and Recurrent Neural Network
Navaneetha Krishnan Rajagopal, Mankeshva Saini, Rosario Huerta-Soto, Rosa Vélchez-Vásquez, J. N. V. R. Swarup Kumar, Shashi Kant Gupta, and Sasikumar Perumal 
Research Article (11 pages), Article ID 5613407, Volume 2022 (2022)

A Massage Area Positioning Algorithm for Intelligent Massage System
Liran Zhou , Zhiquan Feng , Zeyuan Cai, Xiaohui Yang , Changsheng Ai, and Haiyan Shao
Research Article (13 pages), Article ID 7678516, Volume 2022 (2022)

A Novel Fault Diagnosis Method for Denoising Autoencoder Assisted by Digital Twin
Wenan Cai , Qianqian Zhang , and Jie Cui
Research Article (8 pages), Article ID 5077134, Volume 2022 (2022)

MFA: A Smart Glove with Multimodal Intent Sensing Capability
Hongyue Wang, Zhiquan Feng , Jinglan Tian, and Xue Fan
Research Article (15 pages), Article ID 3545850, Volume 2022 (2022)

Deep Multi-Scale Residual Connected Neural Network Model for Intelligent Athlete Balance Control Ability Evaluation

Nannan Xu, Xin Wang , Yangming Xu, Tianyu Zhao , and Xiang Li 
Research Article (11 pages), Article ID 9012709, Volume 2022 (2022)

Potential Future Directions in Optimization of Students' Performance Prediction System

Sadique Ahmad , Mohammed A. El-Affendi , M. Shahid Anwar , and Rizwan Iqbal 
Review Article (26 pages), Article ID 6864955, Volume 2022 (2022)





Intelligent Monitoring System Based on Noise-Assisted Multivariate Empirical Mode Decomposition Feature Extraction and Neural Networks

Le Fa Zhao , Shahin Siahpour , Mohammad Reza Haeri Yazdi , Moosa Ayati , and Tian Yu Zhao 
Research Article (14 pages), Article ID 2698498, Volume 2022 (2022)

Sign Language Recognition for Arabic Alphabets Using Transfer Learning Technique

Mohammed Zakariah , Yousef Ajmi Alotaibi , Deepika Koundal, Yanhui Guo, and Mohammad Mamun Elahi 
Research Article (15 pages), Article ID 4567989, Volume 2022 (2022)

Dynamic and Static Features-Aware Recommendation with Graph Neural Networks

Ninghua Sun , Tao Chen , Longya Ran , and Wenshan Guo 
Research Article (11 pages), Article ID 5484119, Volume 2022 (2022)

Computational Methods for Automated Analysis of Malaria Parasite Using Blood Smear Images: Recent Advances

Shankar Shambhu , Deepika Koundal , Prasenjit Das , Vinh Truong Hoang , Kiet Tran-Trung , and Hamza Turabieh 
Research Article (18 pages), Article ID 3626726, Volume 2022 (2022)





Research on Effect of Load Stimulation Change on Heart Rate Variability of Women Volleyball Athletes

Ludi Liao and Jianying Li 
Research Article (9 pages), Article ID 3917415, Volume 2022 (2022)

A Safe and Compliant Noncontact Interactive Approach for Wheeled Walking Aid Robot

Donghui Zhao , Wei Wang, Moses Chukwuka Okonkwo, Zihao Yang, Junyou Yang , and Houde Liu
Research Article (20 pages), Article ID 3033920, Volume 2022 (2022)

Dynamic Invariant-Specific Representation Fusion Network for Multimodal Sentiment Analysis

Jing He , Haonan Yanga , Changfan Zhang , Hongrun Chen , and Yifu Xua
Research Article (14 pages), Article ID 2105593, Volume 2022 (2022)

Retraction

Retracted: Human Resource Demand Prediction and Configuration Model Based on Grey Wolf Optimization and Recurrent Neural Network

Computational Intelligence and Neuroscience

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

Copyright © 2023 Computational Intelligence and Neuroscience. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] N. K. Rajagopal, M. Saini, R. Huerta-Soto et al., "Human Resource Demand Prediction and Configuration Model Based on Grey Wolf Optimization and Recurrent Neural Network," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 5613407, 11 pages, 2022.

Research Article

AppraisalCloudPCT: A Computational Model of Emotions for Socially Interactive Robots for Autistic Rehabilitation

Ting Yan ¹, Shengzhao Lin ², Jinfeng Wang ³, Fuhao Deng ⁴, Zijian Jiang ⁴,
Gong Chen ⁵, Jionglong Su ⁶, and Jiaming Zhang ²

¹The Brain Cognition and Brain Disease Institute, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong 518055, China

²Institute of Robotics and Intelligent Manufacturing, the Chinese University of Hong Kong (Shenzhen), Shenzhen, Guangdong 518172, China

³Department of Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 215123, China

⁴Shenzhen TOP Intelligent Manufacturing and Technology Co., Ltd., Shenzhen, Guangdong 518129, China

⁵Sunwoda Electronic Co., Ltd., Shiyuan Street, Bao'an District, Shenzhen 518000, China

⁶School of AI and Advanced Computing, XJTLU Entrepreneur College (Taicang), Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 215123, China

Correspondence should be addressed to Jionglong Su; jionglong.su@xjtlu.edu.cn and Jiaming Zhang; zhangjiaming@cuhk.edu.cn

Received 4 August 2022; Revised 13 October 2022; Accepted 21 January 2023; Published 7 March 2023

Academic Editor: Zhongxu Hu

Copyright © 2023 Ting Yan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Computational models of emotions can not only improve the effectiveness and efficiency of human-robot interaction but also coordinate a robot to adapt to its environment better. When designing computational models of emotions for socially interactive robots, especially for robots for people with special needs such as autistic children, one should take into account the social and communicative characteristics of such groups of people. This article presents a novel computational model of emotions called AppraisalCloudPCT that is suitable for socially interactive robots that can be adopted in autistic rehabilitation which, to the best of our knowledge, is the first computational model of emotions built for robots that can satisfy the needs of a special group of people such as autistic children. To begin with, some fundamental and notable computational models of emotions (e.g., OCC, Scherer's appraisal theory, PAD) that have deep and profound influence on building some significant models (e.g., PRESENCE, iGrace, xEmotion) for socially interactive robots are revisited. Then, a comparative assessment between our AppraisalCloudPCT and other five significant models for socially interactive robots is conducted. Great efforts have been made in building our proposed model to meet all of the six criteria for comparison, by adopting the appraisal theories on emotions, perceptual control theory on emotions, a component model view of appraisal models, and cloud robotics. Details of how to implement our model in a socially interactive robot we developed for autistic rehabilitation are also elaborated in this article. Future studies should examine how our model performs in different robots and also in more interactive scenarios.

1. Introduction

Before probing into the scope of computational models of emotions, it is necessary to understand the emotion terminology for the purpose of clarity. To begin with, six terms (i.e., affect, appraisal, cognition, emotion, feeling, and mood) are defined in [1] as follows: (1) to inform one or more cognitive processes, affect which is any information (feeling, mood, and emotion) is used; (2) the process of making

judgments (appraisals) about the relationship between an individual's beliefs, desires, and intentions [2] and perceived events is defined as appraisal; (3) mental processes associated with the comprehension, acquisition, and alteration of knowledge are defined as cognition, such as planning, learning, inference, and recall; (4) owing to concepts and states, emotion is used to inform responses and is defined as cognitive data generated by events (internal and external); (5) the subjective experience of an emotion or of a series of

emotions is defined as feeling; (6) mood is the general state of an emotion that lasts longer and is less variable than the emotion itself.

As for the terminology computational models, according to Simon [3], by drawing results from a system's premises (for example, a weather forecast system), and by predicting the system's behavior, computational models can simulate a system. Furthermore, it was argued that given some premises and appraisal operations, computational models of emotion can predict and potentially produce behavior [4]. According to Marsella et al. [5], computational models of emotions play various roles in research and applications: (1) from the perspective of psychological research, computational models better support our understanding of human emotional processes; (2) from the perspective of AI and robotics research, modeling of emotion can influence the reasoning process or coordinate an agent or robot to better adapt to its environment; (3) from the perspective of HCI research, modeling of emotion improves the efficiency and effectiveness in interaction, as well as enhancing user experience.

To investigate the importance of affective processes in social development and socially situated learning of robots coexisting with humans in the human environment, the computational models of emotions for socially interactive robots were introduced [6]. According to Breazeal et al. [7], to effectively engage in emotion-based interactions, robots must possess three kinds of capabilities: (1) to recognize and interpret human emotional signals, (2) to operate by means of their internal emotional models that are often based on theories in psychology, and (3) to communicate their affective states to others. Since the emotional responses of a robot can be determined by the robot's computational model of emotion which depends on the interaction of its internal cognitive-affective state with the external environment [7], these internal emotional models are crucial for human-robot interactions [8].

Many people with autism spectrum disorder (ASD) have characteristics such as difficulty in social communication (e.g., poor perception of nonverbal cues including facial expressions and gestures in body language, as well as inappropriate expressions), limited and repeated behaviors, as well as narrow, focused interests (in Diagnostic and Statistical Manual of Mental Disorders 5th Edition: DSM 5 [9]). It is increasingly necessary to introduce social interactive robots as an auxiliary means for the treatment and rehabilitation of ASD, so as to improve the diversity of treatment and the effectiveness of rehabilitation training, and to mitigate the medical staff shortages in mainland China and the rest of the world [10]. A number of treatment and training targets, such as triadic interactions, joint attention (JA), turn-taking activities, improving eye contact and self-initiated interactions, assisting the diagnostic process, emotion recognition, and imitation, can be achieved by robotics for autism [11]. Moreover, robots have demonstrated their potentials in 24 of 74 ASD objectives in eight domains, including preschool skills for children with ASD, motor experiences and skills, social/interpersonal interactions and relations, emotional wellbeing, functioning in

daily reality, sensory experiences, and coping, play, and communication [12].

However, better utilization of robots and HCI for autism intervention in the clinical setting does not necessarily lead to robots that are clinically more useful for ASD intervention [13]. This is partly due to the difficulty in the ASD patients to understand the emotional and mental states of others, a feature of the autism spectrum conditions (ASC) [14]. ASC patients show symptoms of stunted development in their ability to recognize and differentiate between different emotional expressions [15]. In addition, children with ASD may be focused on objects of interest for a very long period of time, failing to deliver rehabilitation training outcomes. Hence, if robots are able to follow the gaze, they may be deployed for human-robot interaction tasks, including rehabilitation training for autism [16, 17].

Consequently, when designing computational models of emotions for socially interactive robots, especially for robots for a special group of people such as autistic children, one should take into account the social and communicative characteristics of such a group of people. There are four world-leading research groups with pioneering work in promoting social robots as useful tools in autism therapy, including the Kerstin Dautenhahn Group [18–20]), the Ayanna Howard Group [21, 22], the Maja Matarić Group [23–25], and the Bram Vanderborght Group [26, 27]. However, none of the 4 research groups have designed or applied computational models of emotions for the social robots used in their autism therapy studies. Therefore, this article will propose a novel computational model of emotions that are suitable for and can implement in socially interactive robots, especially for robots adopted in autistic rehabilitation.

The contributions of this article are threefold. First and most importantly, this article presents a novel computational model of emotions so-called AppraisalCloudPCT that is suitable for socially interactive robots that may be used in autistic rehabilitation which, to the best of our knowledge, is the first computational model of emotions built for robots that can satisfy the needs of a special group of people such as autistic children. Second, such a computational model of emotions can enhance human-robot interaction more interactively and effectively, as it takes into account a user's intention and attention and can coordinate the robot to make an appropriate response to the surrounding emotional environment. Third, such a computational model of emotions can achieve a high degree of simulation of human emotions and can be computationally implementable in various robots, as our proposed computational model of emotion is based on the appraisal theories on emotions [28–31], perceptual control theory on emotions [32], a compositional view of model building [5], and cloud robotics [33, 34]/cloud medical robots [35–38].

The rest of the paper is organized as follows. Section 2 revisits some fundamental and notable computational models of emotions that have a deep and profound influence on building some significant computational models of emotions for socially interactive robots, which will be reviewed in Section 3. Section 4 presents our proposed

computational model of robotic emotions so-called AppraisalCloudPCT, and its implementation in a social robot for autistic rehabilitation will be elaborated in Section 5. Finally, the conclusions, limitations, discussion, and future work are given in Section 6.

2. Classical Computational Models of Emotions

The development of computational modeling of emotion and cognition has been accelerated by recent human cognitive and psychological studies related to emotion [1]. For example, according to Marsella et al. [5], concepts drawn from AI have been cast in the appraisal theory of several computational models, including the belief-desire-intention (BDI) model, fuzzy logic, knowledge representation, Q-learning, planning, neural networks, and decision-making.

Marsella et al. [5] used a figure of a “family tree” of a number of the theoretical traditions and significant models (e.g., rational theories, anatomical, dimensional, and appraisal) to illustrate from which they stem. Instead of using a “family tree,” Lin et al. [1] used two tables to review the fundamental theoretical traditions of emotion and cognition and effects modeled by some well-known computational models.

As this article will not focus on the interaction between emotion and cognition as [1] did, cognition theories such as the BDI model will not be reviewed here. Rather, appraisal theories such as OCC (the affect-derivation model proposed by Ortony et al. [39]) and Scherer’s appraisal theory [40], as well as dimensional theories of emotion such as PAD [41], will be revisited here as classical theoretical traditions listed in [5]. Other theories, such as perceptual control theory on emotions [32] and a compositional view of model building [5], which were not listed in the “family tree” in [5], will be also revisited.

2.1. OCC. Ortony et al. [39] proposed an appraisal theory, i.e., the OCC theory in their book “The Cognitive Structure of Emotions,” in which 22 emotions are categorized based on the appraisal of intensity (arousal) and pleasure/displeasure (valence). The OCC theory offers a structure for variables such as the familiarity of an object or the likelihood of an event, to determine the intensity of the emotion types. Based on what is being appraised, the OCC theory broke down the valence appraisal into three categories: praiseworthiness (of an action), like/dislike (of an entity), and desirability (of an event). In addition, when some branches are combined, well-being/attribution compound emotions (e.g., remorse and gratitude) concerning the consequences of events caused by an agent’s actions will be formed.

Specifications have three elements (i.e., the type specification stating the conditions that trigger an emotion of the type, a list of tokens, and a list of variables affecting intensity for each emotion type) that are given for each of the 22 emotion types. The list of tokens specifies which emotional words can be classified as belonging to the type of emotion discussed.

Five negative categories (hate, fear, distress, anger, and disappointment/remorse) and five positive categories (love, relief, hope, joy, gratitude, and pride) from the OCC model

were proposed to use in Ortony [42], in order to decrease the complexity for the development of believable characters. However, for a character using facial expressions only, such ten emotional categories might still be too much, as argued by Bartneck [43], and he proposed to split the emotion process of the OCC model into five phases. Additionally, to resolve the ambiguities identified in the OCC model, a new view of the emotional logic structure of the OCC model based on inheritance was proposed by Steunebrink et al. [44].

2.2. Scherer’s Appraisal Theory. Appraisal theories of emotion, first introduced by Arnold [45] and Lazarus [2, 46], are rooted in Aristotle, Descartes, Spinoza, and Hume [47]. Ellsworth and Scherer and their students actively developed them [28, 40, 48–50] in the early 1980s (see the historical reviews by Scherer [40, 51]). Appraisal theories of emotion relate emotions to the more immediate cognitive assessment of coping capabilities, causal attribution, and evaluation of meaning [52], while the evolutionary theories of emotion relate emotions to biological adaptation in the distant past by contrast. Clore and Ortony [53] treated appraisals to be the psychological representations of emotional significance for the person experiencing the emotion. And Scherer [51] reviewed a central tenet of appraisal theory and arrived at the conclusion that through some dimensions or criteria emotions are triggered and distinguished based on one’s subjective evaluation of personal significance in events, objects, or situations.

Scherer [31] used stimulus evaluation checks (SECs) and defined in the component process model of emotion (CPM) [40, 48, 54–56], to represent the minimum dimension or criteria set sufficient and necessary in distinguishing the essential families of emotional states. The changes in the states of most if not all of the five organismic subsystems will respond to the assessments of external or internal stimuli related to the organism’s primary concerns, and such an episode of interrelated, synchronized changes is defined as emotion [40] in the framework of the CPM (see Figure 1 in [50]). According to CPM, emotion is considered to be a theoretical structure, consisting of five components, each corresponding to one of the five unique functions [50]. In the light of CPM, SECs are processed in sequence of a fixed order, containing four stages in the appraisal process each corresponding to one of the four appraisal objectives, i.e., relevance, implications, coping potential, and normative significance [47]. Moreover, CPM assumes that changes in the internal or external events keep maintaining a recursive appraisal process until the monitoring subsystem sends a signal to terminate or adjust the stimulation triggering the appraisal episode initially [40, 50].

In summary, appraisal theories of emotion not only can be used to investigate the origin of emotion but can also be used to account for the emotions of people experiencing feelings, using the Geneva Emotion Wheel (see the second version in [57]), or the Geneva Expert System on Emotion (<https://www.unige.ch/cisa/properemo/gep17/intro1.php>). In addition, facial expressions and physiological processes

may change during the evaluation or appraisal of the personal significance of a certain object or situation, but which discrete emotion is experienced can be determined by the specific profile of appraisal (i.e., the antecedent of the emotion), according to Niedenthal et al. [52]. As a result, two individuals can experience different emotions despite being subjected to the same event or stimulus, which is consistent with the appraisal theories of emotion.

2.3. PAD. According to dimensional theories of emotion, emotion and other affective phenomena should be classified and labeled in the way of the social construction—as points in continuous (usually two- or three-dimensional) space but not as discrete entities [41, 58–60]. The historical development of dimensional theories of emotion can be traced back to James [61], Schachter and Singer [62], Russell [58], and Barrett [59]. Russell [63] suggested replacing discrete emotions with core affect due to cross-cultural differences which attribute specific emotions to facial expressions. Scarantino [64] described the “core affect” as follows:

“Core affect, understood as the category comprising the set of all possible valence and arousal combinations on the circumplex, differs from discrete emotions in three crucial ways: it is ubiquitous, it is objectless, and it is primitive.” ([64], p. 948).

According to Russell ([58], p. 154), a person is in exactly one affective state at any time and such possible core affective states can be characterized in the space of continuous and broad dimensions. Mehrabian and Russell’s “PAD” model [41] consists of three dimensions corresponding to pleasure (measuring valence), arousal (to measure the level of affective activation), and dominance (a measure of control or power), respectively. Many computational models of emotion were inspired by the PAD model, such as WASABI [65], a PAD-based model of core affects incorporating Scherer’s sequential-checking theory.

2.4. Perceptual Control Theory on Emotions. Perceptual control theory (PCT) [66] is a theory on how living organisms can control their inputs instead of their outputs. The idea of PCT can be attributed to [67]: “What we have is a circuit, not an arc or broken segment of a circle. This circuit is more truly termed organic than reflex because the motor response determines the stimulus, just as truly as sensory stimulus determines the movement ([67]; p. 363).” “PCT was developed by William T. Powers, a physicist/engineer, in the 1950s. He first published it in [68], then formalized it in [66], and revised it in his latest work [32]. According to PCT, through some principles, behavior is defined as (merely) the control of perception: (1) negative feedback leads to control; (2) a specific hierarchical organization of loops leads to control; (3) perception can be only controlled by individuals themselves; (4) conflicts can be caused by controlling others; (5) “dysfunction” can be caused by conflicts between high-level control systems; (6) a specific learning mechanism helps reorganization reestablishes control.”

Moreover, PCT states that control systems are organized in a hierarchy to manage complex goals such as controlling

low-level motor as well as regulating high-level psychological and social behavior, by defining the reference signal for the layer below in each layer [66]. The levels of hierarchical perceptual control theory hypothesized by Powers are, respectively, 1st-order: intensity; 2nd order: sensation/vector; 3rd-order: configuration; 4th-order: transitions; 5th-order: sequence; 6th-order: relationships; 7th-order: program; 8th-order: principles; 9th-order: system concepts.

In addition, Powers explained how to generate emotions through a PCT model in his paper [69]: (1) as the brain regulates, the neurochemical reference signals sent from the hypothalamus through the pituitary gland to all major organ systems, and emotion is defined as a product of brain activity; (2) as perceivable changes of physiological state result from disturbances calling control systems into action, emotion is a direct response to the disturbance, the presence of which can be known of instantly by one’s conscious awareness; (3) in closed-loop terms, an experienced emotion is caused by “feelings” which is a collection of inputs and perceptions; meanwhile, it outputs a change in the physiological state (e.g., vasoconstriction, respiration rate, metabolism, heart rate, and motor preparedness); (4) an emotion is caused to happen by a reference signal in some high-level system specifying more or less intended amount of some perception, but not by the external factors; (5) in a high-level control system, a zero error signal results from that the perceived current state matches the specified reference signal; while the mismatch will cause a nonzero error signal, so action needs to be taken to correct the error causing emotion; (6) emotional behavior and emotional thinking can be caused by an error signal immediately resulting from a change of reference signal or a change of a disturbance; (7) the strongest negative emotions are related to the largest errors and errors that human beings think need to be corrected most, and when some internal or external factors prevent us from taking action to correct errors, their maximum intensity and duration will appear; (8) when the degree of error is significant and important to them, human beings will use emotional words, leading to awareness of the cause, while small errors mean not using emotional words, leading to failure to identify the cause.

To summarize, emotions are defined to be one aspect of the wholly integrated hierarchy of control by PCT on emotions. The PCT on emotions involves the notion of “an embodiment” (e.g., emotion is defined as a product of brain activity), “adaptation” (e.g., the “general adaptation syndrome” in the case of attack behavior or avoidance), and “appraisals” (e.g., evaluating the significance of an error signal). Consequently, PCT on emotions is compatible with other theories such as the theory of embodied emotion, evolutionary theories, and cognitive-appraisal theories to some extent.

2.5. A General Architecture of Computational Models of Emotion. Marsella et al. [5] argued that a number of component “submodels” integrated into the computational models listed in the “family tree” are not clearly delineated. They proposed that by disassembling “submodels” along

appropriate joints, a large number of significant differences between different computational models of emotion can be decomposed into a few design choices.

They then proposed a component model view of appraisal models conceptualizing emotions as a set of linked component models (see Figure 2 in [5]) and the relationships between these components. Terminology associated with each of the component models listed in the appraisal architecture was also introduced: (1) person-environment relationship: the term refers to some expression of the relationship between the agent and its environment, which was introduced by Lazarus [2]; (2) appraisal-derivation model: such a model converts some representations of the relationship between a person and the environment into a set of appraisal variables; (3) appraisal variables: they are a set of specific judgments generated as a result of an appraisal-derivation model, which can be used by an agent to produce different emotional responses; (4) affect-derivation model: the mapping from appraisal variables to affective state is processed in this model, and once a pattern of appraisals has been determined, then accordingly how an individual will react emotionally is also specified in this model; (5) affect-intensity model: in the model, a specific appraisal will result in the strength of the emotional response, which is usually calculated by an intensity equation using a subset of appraisal variables, such as desirability and likelihood; (6) emotion/affect: affect could be a set of discrete emotions, a discrete emotion label, core affect in a continuous dimensional space, or even a combination of these factors; (7) affect-consequent model: this model maps affect (or its antecedents) onto some behavioral or cognitive changes which are determined by the behavior consequent models and cognitive consequent models, respectively. Behavior consequent models summarize how affect (e.g., emotion, feeling, and mood) alters an agent's observable physical behavior such as facial expressions, while cognitive consequent models determine how affect will change the nature or content of cognitive processes such as an agent's beliefs, desires, and intentions, respectively.

Three rather different systems, i.e., EMA [70], ALMA [71], and FLAME [72], were characterized in [5] to highlight the conceptual similarities and differences between emotion models by using a component model view of appraisal models. Marsella et al. [5] argued that the adoption of a component view of the model building can empirically assess the capabilities or validity of alternative algorithms to implement the model and conduct meaningful comparisons (i.e., similarities and differences) between systems.

To sum up, Marsella et al.'s compositional view of model building [5], which lays stress on that emotional models, is often composed of individual "submodels" or "smaller components" that can be matched, mixed, or excluded from any given implementation and is often shared. According to Marsella et al. [5], components may be evaluated and subsequently abandoned or improved due to ongoing evaluations before the final version of the model is designed.

3. Classical Computational Models of Emotions for Socially Interactive Robots

3.1. Kismet's Cognitive-Affective Architecture. With four perceptual modalities (facial display, body posture, gaze control, and speech), an expressive robot called Kismet [73] was developed by MIT, to explore the nature of social interaction and communication between humans. In other words, insights from psychology and ethology [8] have inspired the extensive computational modeling, to explore the social interaction between caregiver and infant.

In view of the key role of infants in normal social development, in order to implement core primitive social response shown by infants, a cognitive-affective architecture emphasizing interactive and parallel systems of cognition and emotion [6] was designed for Kismet. The architecture (see Figure 58.6 in [8]) mainly contains two parts, one is the cognitive systems which are responsible for drives, attention, perception, and goal arbitration while the other part is the affective processes that include affective appraising incoming events, expressive motor behavior (facial expressions, vocalizations, etc.), and basic emotive responses. Therefore, Kismet's models of emotion interact closely with its cognitive system, affecting the behavior and goal arbitration in the architecture [7].

By combining the basis facial postures, Kismet produces a continuous range of expressions (i.e., five primary emotions (happiness, fear, disgust, sadness, and anger) and three additional ones (excitement, interest, and surprise) of varying intensities. This is achieved through the application of an interpolation-based technique in a three-dimensional, componential affect space consisting of the valence, arousal, and stance axes [74], adapted from Russell's circumplex model (arousal and valence) [75], and resonated well with the work of Smith and Scott [76]. Breazeal [74] enumerated a number of advantages gaining from this affect space, such as making the reception of robot facial expressions clearer since only a single state can be expressed at a time (according to selection), enabling reflecting the nuances of the underlying assessment of the robot's facial expressions, and facilitating smooth trajectories through the affect space.

The importance of building an emotional space that allows smooth transitions between discrete emotions was emphasized by the Kismet project, although it does not compare the believability of the expression of smooth transitions and nonsmooth transitions. Moreover, the Kismet project shows that by using a computational model of emotion, a robot can conduct social interaction with humans apart from arbitrating its internal affective states [77, 78].

3.2. WE-4RII's Mental Model. The core of the mental model of a robot called WE-4RII (see Figure 58.10 in [8]) is the emotion model. The dynamics of mental transitions in the WE-4RII mental model can be expressed by equations adopting the equation of motion that describes the movement of objects in dynamics [8].

To express the dynamics of mental transitions, the WE-4RII robot has implemented equations of emotion, mood vector, and equations of need (see [79] for more details). Furthermore, the seven basic emotions defined by Ekman [80] are represented as the emotion vector [79, 81] in a three-dimensional mental space consisting of the pleasantness, activation, and certainty axes. Seven emotions and the expressions corresponding to these seven emotions are mapped into a 3-D mental space, and the regional mapping of WE-4RII's emotions is determined by the emotion vector E passing through each region (see Figure 58.11 in [8]).

In summary, the mental model of WE-4RII can be computationally implemented [82], as it implements equations inspired by motion to express the dynamics of mental transitions.

3.3. PCT-Based Model PRESENCE. To generate robotic emotional behavior, some researchers have designed some computational structures based on PCT on emotions. For instance, a model called PRESENCE “PREdictive SENsor-intor Control and Emulation” which is based on PCT was developed by Moore [83] to improve the speech-based human-machine interaction. Due to PRESENCE, a system can cater to the needs and attention of a user, while a user can allow for the needs and intentions of the system. According to Moore [83], cooperative and communication behaviors are by-products of recursive hierarchical feedback control structures based on this ensemble model.

Some theories and ideas in domains, such as control, neuroscience, bioscience, and psychology, have laid a foundation for the creation of PRESENCE. These theories and ideas include “perceptual control theory,” [66] “mirror neurons,” [84] “hierarchical temporal memory,” [85] and “emulation mechanisms.” [86] To solve three fundamental constraints (i.e., energy, entropy, and time) that ultimately determine the organism's ability to survive within an evolutionary framework, PRESENCE was originally designed as an integrated and recursive processing architecture. To facilitate efficient behavior and efficient communications, PRESENCE maximizes the achievements of the system or the user in the interactive environment, and it is organized into four layers and is therefore inherently recursively nested and therefore hierarchical in structure.

The PRESENCE has been demonstrated in [83] that a Lego NXT computer model was built by Moore to maximize the synchronization of its own behavior with external sources. The robot can sense external sources such as external sounds, can sense its own sounds, and can generate its own rhythmic behavior. Moore's research shows that PCT not only can be used for explaining emotional behavior but also be used in the prediction of emotional behavior.

3.4. iGrace Computational Model of Emotions. The iGrace computational model (see Figure 1 in [87] for more details) was designed to enable a companion robot EmI to have a nonverbal emotional response to the speaker's speech. The iGrace consists of 3 principal parts, i.e., the “input” module, the “emotional interaction” module, and the “expression of

emotions” module, which can enable EmI to receive input information, process them, and determine emotional behavior. Saint-Aimé et al. [87] described these three modules as follows.

The 7 uplets of the understanding module (i.e., the act of language, actions “for the child,” concepts “for the child,” tense, coherence, phase, and emotional state), the audio signal, and the video signal are taken into account in the “input” module. As such, this module can represent the interface for data exchange and communication between the emotional interaction module and the understanding module.

With the “emotional interaction” module, iGrace can generate the emotional state of EmI using discourse information given by “input” as well as its internal cognitive state. This module contains 4 submodules, namely moderator, selector of emotional experience, generator of emotional experience, and behavior (see more details in [88, 89]), which produce lists Li of pairs (eemo, C (eemo)) involving in four steps (see Figure 2 in [87]) in which C (eemo) denotes an influence coefficient and eemo denotes an emotional experience.

In the “expression of emotions” module, a list of triplet < tone, posture, facial state > is built to express the emotional state of EmI, in which tone is converted into music notes and postures, and facial expressions of EmI are converted into motor movements.

To sum up, the iGrace computational model has demonstrated that it can be computationally implemented in companion robots such as EmI and the new version of EmI [87]. This might result from that iGrace is an instance of the generic model of emotions GRACE [90]. Furthermore, as compared to other computational models of emotions, such as FLAME [91], Kismet [7], Greta [92], EMA [70], and GALAAD [93], GRACE is the only model that applies the three fundamental theories that characterize an emotional process, namely the appraisal theory, coping theory, and personality theory, according to Saint-Aimé et al. [88].

3.5. A Computational System of Emotion xEmotion. To allow an agent (a robot carrier) to respond most appropriately to specific changes in the environment, xEmotion, a computational system of emotion, is designed. According to Kowalczyk et al. [94], implementing the intelligent system of decision-making (ISD) in an autonomous agent or robot can make it operate faster and more efficiently, resulting from the ISD's system of emotions, which can be viewed as an approach based on scheduling variable policies from a control theory perspective. Covering various psychological theories on emotions such as the somatic, evolutionary, and appraisal theories of emotion, xEmotion takes into account specific temporal divisions of emotion and, in particular, considers both long-term changes (e.g., personality changes or emotional disorder) and short-term emotions (e.g., expressions or autonomous changes). Furthermore, xEmotion uses (common/real and private/imaginary/individual) wheels/circles of emotion or the “rainbows” of emotions [95] to interpret and compile emotions.

Kowalczyk et al. use a general scheme (see Figure 3 in [94]) to explain how emotions are used as a scheduling variable in the xEmotion system. It takes approximately five big steps to generate emotions in the scheme, namely impression recognition, discoveries recognition, generating emotion/generating equalia (these two phases are parallel), generating mood, and available reactions. And 6 principal components of xEmotion, i.e., autonomous preemotions, expressive subemotions, expressive subequalia, classic emotion, equalia, or private emotion, and mood are distinguished in [94, 96, 97].

For xEmotion to be computationally implementable in the agent (robotic carrier), Kowalczyk et al. [94] applied fuzzy sets in six principal components of xEmotion in the three phases of an emotion process, namely somatic emotions (or preemotions), appraisal of emotions (including subemotions and emotion), and personal emotions (including subequalia, equalia, and mood). The emotional components and their underlying relationships can be found in [98].

In summary, emotions in xEmotion are used not only as scheduling variables (for decision-making and forming responses or general behavior) but also as adjustment parameters (in the motivation subsystem). Furthermore, interpreting and using emotions as a scheduling control variable have made some contributions to the research and the implementation of the computational model of emotions for robots.

4. Our Proposed Computational Model of Robotic Emotions

In this section, we first propose a computational model of emotions for socially interactive robots, especially for robots for a special group of people such as autistic children, so-called AppraisalCloudPCT (based on a component view of computational models, the appraisal theories on emotions, cloud robotics, and perceptual control theory on emotions), then we compare our model AppraisalCloudPCT with the five models for robotic emotions revisited in Section 3.

4.1. Our Computational Model AppraisalCloud PCT

4.1.1. Principles in Design of a New Model. There are certain key points we want to stress, or several problems (e.g., how can a robot highly emulate human emotions? how can a computational model of emotions be highly computable? how can a computational model of emotions be suitable for most of the socially interactive robots and be computationally implementable in them?) we want to tackle when designing a new computational model of robotic emotions.

The followings are given five primary principles in designing our new model:

- (1) Principle in the simulation of human emotions: with a computational model of emotions, a robot may simulate the whole process of a human emotion (e.g., generation, regulation, and responding to a stimulus), and as such, the computational model can highly simulate a human emotion

- (2) Principle in achieving computability of the model: each component of the model should be computable, and as a whole, the model should be computable
- (3) Principle in enhancing human-robot interaction: a computational model of emotions should not only take into account a user's intention (or need), attention, emotional state, the response to the robot, and the impact of the external environment (such as noise, disturbance, contextual cues) on the user during the interaction but also coordinate the robot to make an appropriate response to the surrounding emotional environment
- (4) Principle in promoting the universality of a computational model in socially interactive robots: as more and more socially interactive robots are deployed in therapy and rehabilitation situations, a computational model of emotions should take into account the social and communicative characteristics of a special group of users such as autistic children or dementia elders
- (5) Principle in facilitating sharing information between and learning from socially interactive robots: a computational model of emotions should endow a robot with a more powerful capability of making decisions faster, more appropriate, and more efficient, given that more and more socially interactive robots will be exposed to various users with different backgrounds and be connected to substantial Internet of Things (IoT) such as medical IoT with massive medical data

4.1.2. An Overview of the New Model Appraisal Cloud PCT. Based on the five primary principles in designing a model mentioned previously, we designed a new computational model of robotic emotions AppraisalCloudPCT as illustrated in Figure 1.

The theoretical basis and guiding methodology covered in the proposed model in response to each of the 5 primary principles are introduced as follows:

- (1) The proposed computational model adopts the concepts of perceptual control theory (PCT) on emotions [32] and PCT-based PRESENCE [83] to achieve simulation of human emotions: in a closed-loop as illustrated in Figure 1, a collection of the intention of a robot (i.e., a reference signal) and achievement (perceived outcome) (i.e., a perceptual signal) will cause an experienced emotion, and at the same time, an output-caused change in the cognitive states and behavior of the robot will affect a user's behavior during the human-robot interaction. In other words, the difference (i.e., a mismatch) between the reference signal and the perceptual signal will immediately result in an error signal, which will give rise both to the emotional behavior and to the emotional thinking of a robot. And emotions with greater intensity and longer duration will arise in connection with a larger error that demands a robot

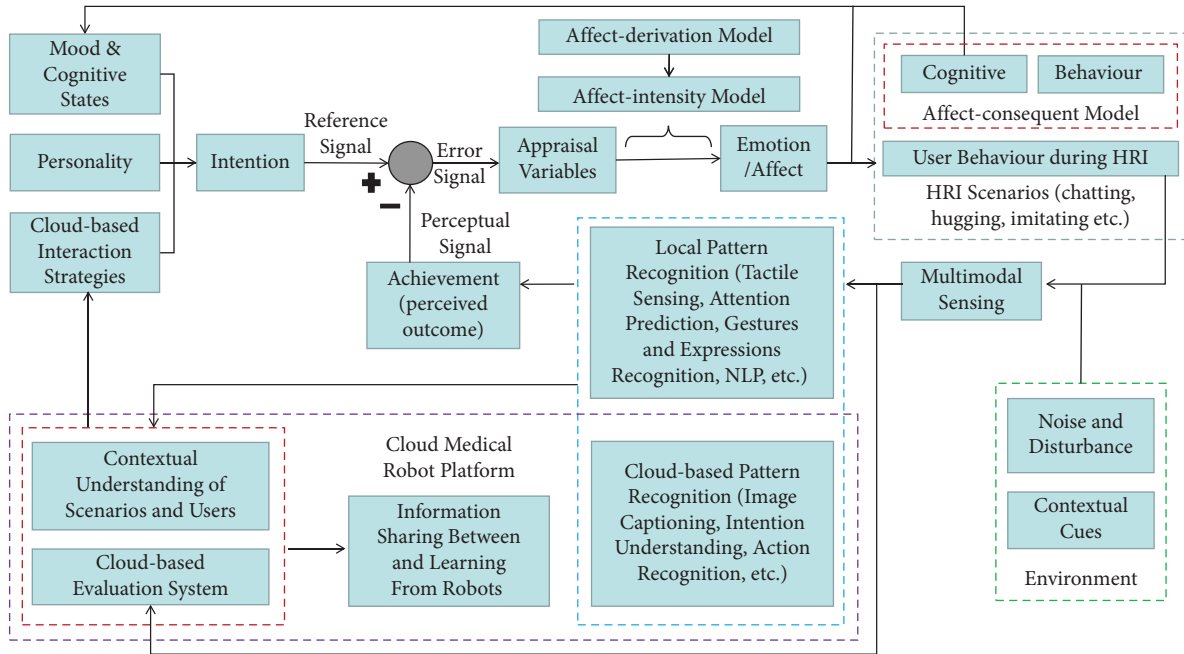


FIGURE 1: The whole architecture of the proposed model AppraisalCloudPCT.

to alter its affect-consequent model more appropriately to correct the error. Moreover, with the computational model, a robot will be endowed with mood and cognitive states, personality, and cloud-based interaction strategies to form its intention. As such, the computational model can highly simulate the whole process (e.g., intention, generation, regulation, and responding to a stimulus) of a human emotion.

- (2) The proposed computational model adopts Marsella et al.'s compositional view of model building [5], which lays stress on that emotional models are often composed of individual "submodels" or "smaller components" that can be matched, mixed, or excluded from any given implementation and are often shared. And 5 out of 7 component models listed in the appraisal architecture in [5] are adopted in our proposed computational model, namely appraisal variables, affect-derivation model, affect-intensity model, emotion/affect, affect-consequent model consisting of the cognitive-consequent model, and behavior-consequent model. As illustrated in Figure 1, our proposed model is assembled from more than 15 "submodels." Consequently, when each of them is computable, the computability of our proposed model as a whole can be achieved.
- (3) On one hand, to make human-robot interaction more effective, efficient, or pleasant, the achievement (perceived outcome), e.g., the interpretation of a user's intention, attention, emotional state, and behavior, will influence the appraisal variables in the proposed computational model. On the other hand, to coordinate a robot to respond to the surrounding emotional contexts (i.e., contextual cues in the

environment containing emotional information that might have an impact on a user's interpretation of the behavior of a robot [99–102]) appropriately, so that the robot can fit with its environment better, contextual understanding of the scenarios and the user is taken into account to support the cloud-based interaction strategies in the proposed computational model.

- (4) The proposed computational model of emotions takes into account the social and communicative characteristics of a special group of users such as autistic children or dementia elders, through its submodel so-called cloud-based interaction strategies, which is supported by two submodels (i.e., "contextual understanding of scenarios and users" and "cloud-based evaluation system") in a cloud medical robot platform, as illustrated in Figure 1. A cloud-based evaluation system may have certain advantages as mentioned in the research on cloud medical robots [35–38], one of which is data of interaction between a user and a robot can be stored and evaluated in a cloud for further assessment of the social and communicative characteristics of a user. And the submodel "contextual understanding of scenarios and users" relies on another two submodels "local pattern recognition" and "cloud-based pattern recognition," which can provide the interpretation of a user's intention, attention, emotional state, and behavior. Therefore, the proposed computational model of emotions is suitable for socially interactive robots, especially for robots for a special group of users such as autistic children or dementia elders, which promotes the universality of our model to some extent.

- (5) To facilitate sharing information between and learning from socially interactive robots, a cloud medical robot platform is built and assembled in the proposed computational model. With such a platform, information can be shared between robots through the submodel “cloud-based evaluation system,” and the capability of interpretation of a user and of making decisions can be learned through the submodel “contextual understanding of scenarios and users.”

4.2. Comparison of Models. This section compares the 5 computational models for robotic emotions revisited in Section 3 with our proposed computational model (see Table 1 for a summary). The five crucial properties of a computational emotion model, (i) domain-independent, (ii) models mood, (iii) models personality, (iv) data-driven mapping, and (v) ethical reasoning, as listed in a review paper [103], alongside with one more property (vi) combining with cloud robotics (we believe this will be a future trend in building the computational models of emotions for socially interactive robots), are chosen as the six criteria for comparison.

Table 1 shows a comparative assessment between the computational models of emotions for socially interactive robots as can be inferred from the summary in the table, even to satisfy the first five criteria still remains as a challenge. Great efforts have been made in building our proposed computational model of emotions to meet all the six criteria, by adopting the appraisal theories on emotions, perceptual control theory on emotions, a component model view of appraisal models, and cloud robotics. How our proposed computational model meet all the six criteria is summarized as follows: (1) to meet of the criteria of “domain-independent,” our proposed computational model not only takes into account the social and communicative characteristics of every user but also can coordinate a robot implementing our model to respond to the surrounding emotional contexts appropriately; (2) mood is considered as a long-term change in a submodel “mood and cognitive states” of our proposed computational model, and it is impacted by the other two submodels “emotion/affect” and “cognitive,” and therefore, the second criteria “models mood” can be met; (3) there is a submodel “personality” in our proposed computational model such that personality can be modeled; (4) between appraisal variables and emotions, there are two consecutive submodels “affect-derivation model” and “affect-intensity model” in our proposed computational model, which supports data-driven mapping of the appraisal variables into emotion intensities; (5) a emotion regulation mechanism is implemented in our proposed computational model through a closed-loop emotion modeling and regulation based on perceptual control theory on emotions, and through a submodel “cloud-based interaction strategies,” (6) our proposed computational model combines with cloud robotics by using a submodel “cloud medical robot platform.”

5. The Implementation of Our Model in a Social Robot for Autistic Rehabilitation

5.1. A Social Robot for Autistic Rehabilitation. We developed a socially interactive robot so-called Dabao for autistic rehabilitation, with which we conducted three preliminary clinical human-robot interaction studies [10, 104, 105] for Chinese children with ASD. The appearance and functionalities of the robot are demonstrated in Figure 2, and the software architecture is illustrated in Figure 3 as follows.

Apart from the tactile sensing [106] and some APP instances [105, 107, 108] on the touch screen as demonstrated in Figure 2, we have developed some other deep learning algorithms to endow the robot with a stronger capability in the interpretation of a user (e.g., an autistic child), such as intention understanding (see Figure 4) and attention recognition (see Figure 5). Furthermore, Table 2 summarizes six major capabilities of the robot to perceive a user, to infer a user’s mood and cognitive states and behavior, and to express itself to a user that can influence the effect, the efficiency, and the pleasantness in the human-robot interaction.

5.2. The Implementation of Our Model in the Social Robot. As illustrated in Figure 6 (as equivalent to Figure 1, except for all of the submodels are marked in different numeric symbols and different color themes, for a better explanation of how our model is implemented in the social robot Dabao developed by us), our proposed model AppraisalCloudPCT consists of 20 compositional submodels (or so-called components of a model). Such a compositional view of the model building has certain advantages, one of which is that we can implement the proposed model AppraisalCloudPCT in our social robot by implementing its compositional submodels one by one and then by forming the whole model in a closed-loop control.

We implement each submodel with mathematical definitions and formulas in our social robot as follows.

5.2.1. The 1st Submodel “Mood and Cognitive States”. The equation of mood, equation of emotion, as defined in [112], will be adopted in implementing our proposed model AppraisalCloudPCT. First, emotion vector E can be defined in the PAD mental space consisting of the pleasantness, arousal, and dominance axes as the robot’s cognitive state as follows:

$$E = (E_p, E_a, E_d), \quad (1)$$

where E_p is the pleasantness component of emotion, E_a is the arousal component of emotion, and E_d is the dominance component of emotion.

According to Itoh et al. [112], mood vector M , consisting of a pleasantness component and an arousal component, can be defined as follows:

TABLE 1: Comparison of six computational models of emotions for socially interactive robots.

Models	Criteria					
	Domain-independent	Models mood	Models personality	Data-driven mapping	Ethical reasoning	Combines with cloud robotics
Kismet	✓	?	×	✓	✓	×
WE-4RII	✓	✓	✓	×	?	×
PRESENCE	✓	×	×	✓	✓	×
iGrace	✓	✓	✓	✓	✓	×
xEmotion	✓	✓	✓	✓	?	×
Our model	✓	✓	✓	✓	✓	✓

Note. (1) A model that satisfies the given property is marked with a tick mark (✓); a model that does not satisfy the given property is marked with a cross mark (×), and when we were unable to retrieve enough information to determine whether a specific property was met, we use a question mark (?). (2) According to Ojha et al. [103], “domain-independent” means processing and exhibiting emotional responses in various situations but not only in certain kinds of interaction domain; “models mood” means integrating the notion of mood with emotions; “models personality” means integrating the notion of personality; “data-driven mapping” is defined as a data-driven mapping of the appraisal variables into emotion intensities according to the learned relationship between emotions and appraisal variables; as for “ethical reasoning,” it is defined to be an emotion regulation mechanism implemented based on ethical reasoning for the emotional and behavioral responses of social robots to be more “acceptable” in the human community.

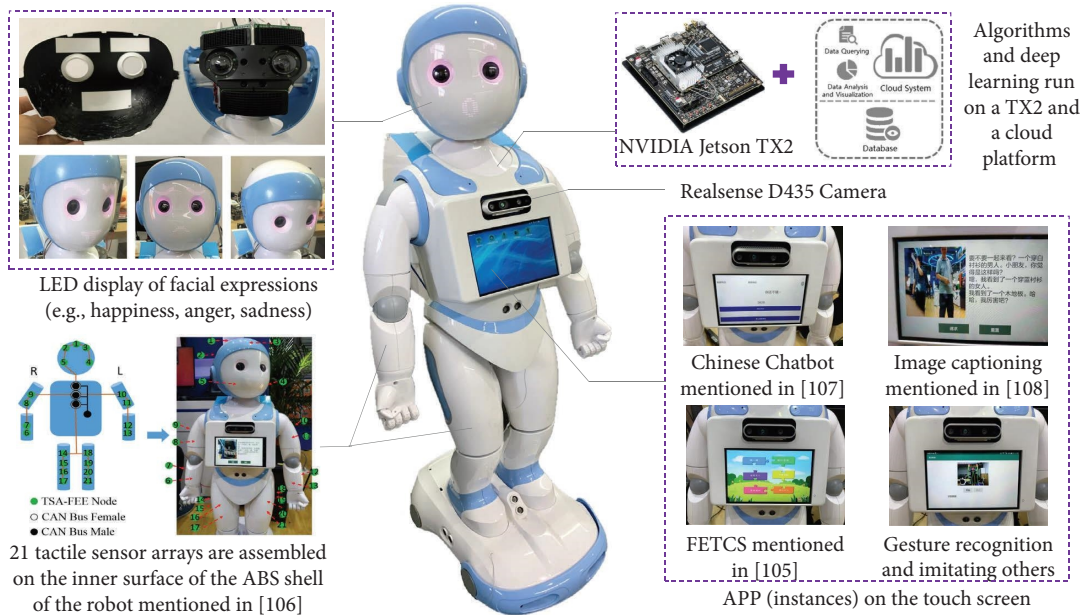


FIGURE 2: The appearance and functionalities of our developed robot Dabao for autistic rehabilitation.

$$M = (M_p, M_a, 0), \quad (2)$$

$$M_p = \int E_p dt, \quad (3)$$

$$\ddot{M}_a + (1 - M_a^2)\dot{M}_a + M_a, \quad (4)$$

where M_p and M_a denote the pleasantness and arousal components of the mood, respectively. The integral of the pleasantness component of the emotion equation (3) defines M_p , resulting from that the pleasantness of mood can be influenced by the current cognitive state. Furthermore, M_a has been defined by the Van del Pol equation (4) owing to that the activation component of mood vector is similar to the biological rhythm of the human body, such as the internal clock.

5.2.2. *The 2nd Submodel “Personality”*. By far, the big five personality traits (i.e., openness (O), conscientiousness (C), extraversion (E), agreeableness (A), and neuroticism (N)), as defined in [113, 114]), were the most widely used measure for human and robot personality modeling in human-robot interaction literature. Three main conclusions can be drawn from the literature review in [115]: (1) extroverts seemingly react more positively in the period of interaction with robots; (2) humans respond more positively to extroverted robots, but this relationship is moderate; (3) humans respond well to robots with similar and/or different personalities. Furthermore, Robert [115] suggested the effects of context on the impact of robot and human personality to be looked at in future studies, as it is easy speculating that the personality of a robot may be more important to a home robot rather than one used at work. This is consistent with the contextual approach to personality, whereby a person’s personality is

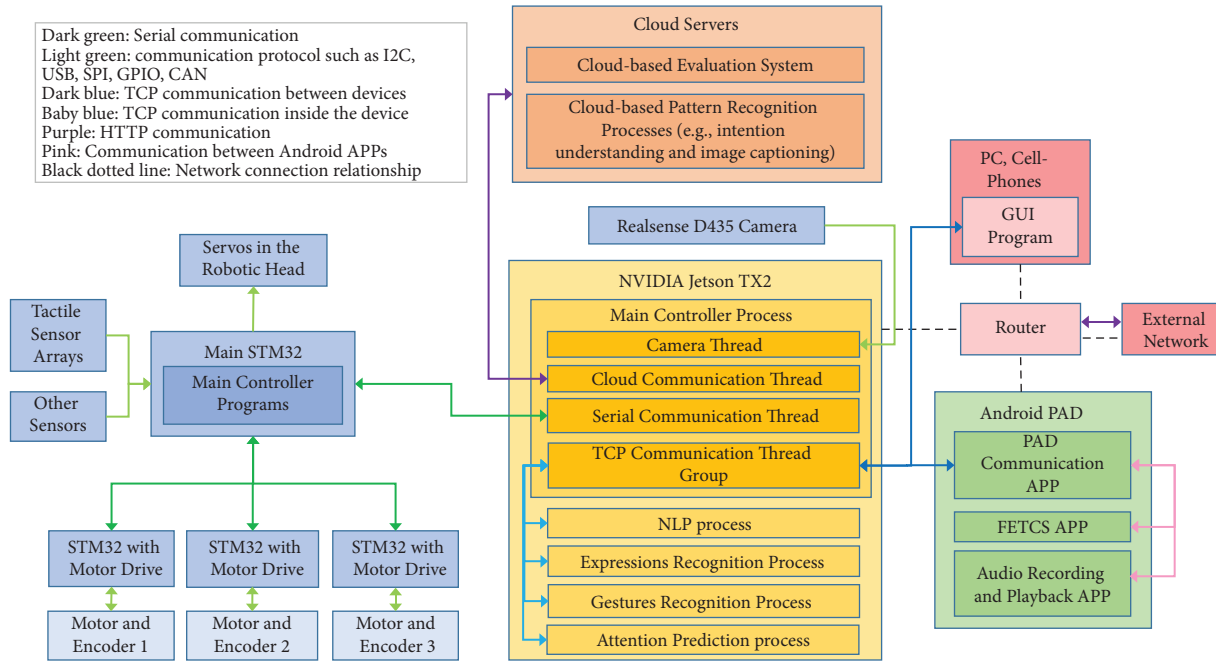


FIGURE 3: The software architecture of the robot.

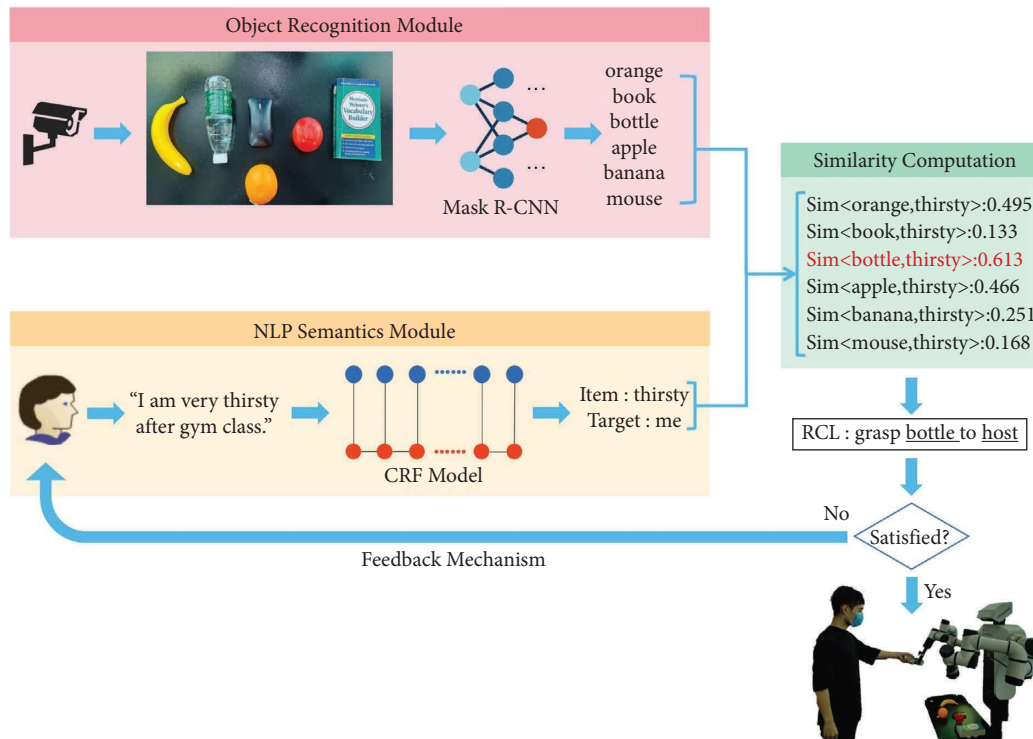


FIGURE 4: A new task-based framework that enables robots to understand human intentions using visual-NLP semantic information [109]: it includes a language semantics module to extract keywords no matter if the command directive is explicit or not, a visual object recognition module to identify multiple objects located to the front of the robot, and a similarity computation algorithm for inferring the intention based on a given task (i.e., selecting some desired item out of multiple objects on a table and giving it to a particular user among several human participants). Result of the similarity computation is then translated into structured robot control language RCL (grasp object to place) to be comprehended by robots. The experimental results demonstrate the ability of the framework to allow robots to grasp objects with the actual intent of vague, feeling, and clear type instructions.

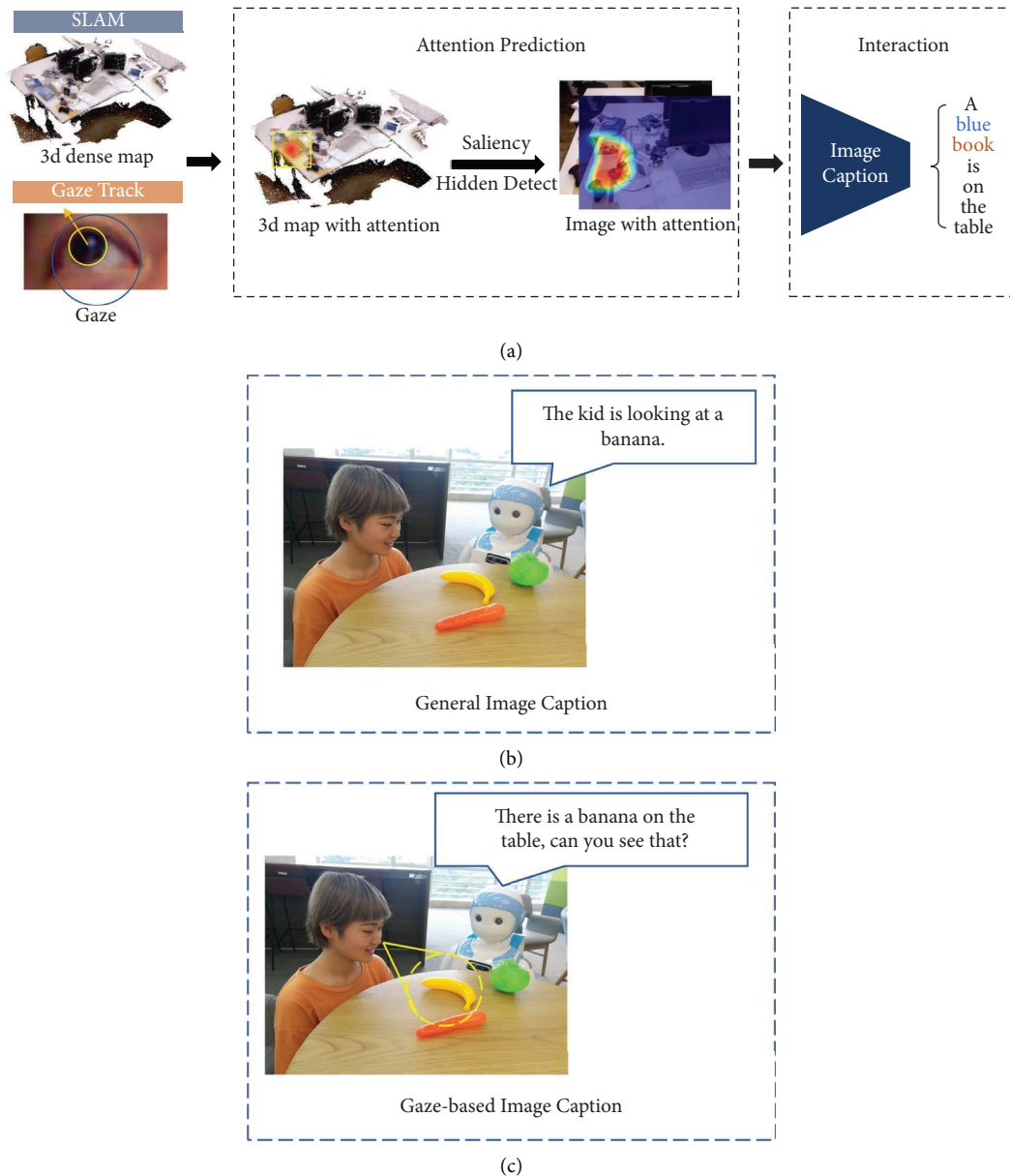


FIGURE 5: The overall framework of a novel gaze-based image caption system for autistic children and the effect of the framework in a gaze-based image caption system [110]: (a) the overall framework describes the region where an autistic child is looking at and combines image caption (based on attention heat maps, it describes the region concentrated by the child) with gaze-following (it is based on spatial geometry and predicts areas of attention from the spatial relationship between the map and line of sight); (c) is more suitable than (b) in enhancing human-robot interaction and in promoting the spontaneous language development of autistic children, as adding gaze-following can support a robot in better describing what the child is looking at.

best described and understood in the various contexts in which it is placed [116]. Moreover, the users' preferences for robot personalities can be determined by people's stereotype perceptions of certain jobs and the background of the robot's role [117]. Therefore, the behavior of the robot may need to be adapted to the user's expectations as to what personality and behavior are consistent with such tasks or roles.

In a recent study [118], researchers found that participants performed better when using a robotic assistant with a similar personality to their own or a human assistant with a different personality. This is in accordance with the results

of the systematic evaluation of human and robot personality in healthcare human-robot interaction [119] that matching the patient and robot personality based on introversion or extroversion is positively correlated with beneficial results. Research in [119] also found that robot personality traits such as extroverted, feminine, responsive, amiability, and sociable were positively associated with beneficial outcomes.

Not only the emotional factors [120] but also the appraisal patterns of emotion [121] can be affected by the Big Five personality traits. The relationship between the PAD model [41] and the five factors of personality can be derived

TABLE 2: A summary of the six key capabilities of our developed robot Dabao in interactive scenarios with Chinese autistic children.

Interactive scenarios	Perception (i.e., to perceive an autistic child via cameras, tactile sensor arrays, microphones, etc.)	Capabilities	Action (i.e., to express itself to an autistic child through facial expressions, gestures, speech, etc.)
Recognizing the gestures of an autistic child and imitating the user	The robot can recognize five hand gestures (e.g., OK, punch) and ten body gestures (e.g., kick, wave)	There is a mapping between the skeleton feature key points of an autistic child and the output gestures of the robot	The robot can either imitate the gestures of the autistic child or feedback to the child with one of the seven emotional body gestures accordingly
Chatting with an autistic child [107]	Via automatic speech recognition (ASR) The robot can sense four types of primary tactile characteristics (i.e., area, location time, and direction) of an autistic child via 21 tactile sensor arrays by our own design [106]	Via natural language processing (NLP) Ten machine learning algorithms including SVM and KNN are selected using K-fold cross-validation to classify the 6 touch behaviors (e.g. finger sliding) of the child	Via text-to-speech (TTS) There is a mapping between an event-triggered tactile perception and the output (facial expressions, gestures, speech, etc.) of the robot
Recognizing the facial expressions of an autistic child and responding to the child accordingly	Via our lightweight CNN architecture DeepLook [105] or Concat-Xception [111] that runs on Jecton TX2, our robot can recognize the 6 basic facial expressions (e.g., anger) of a child at an average rate over 70%	There is a mapping between the facial expressions of a child perceived by the robot and the output facial expressions of the robot that takes into account the child's ability to perceive	An appropriated facial expression out of 20 choices (shyness, thinking (turning eyes), etc.) will be displayed to the child with body gestures and speech output sometimes
Attention-based image captioning with gaze-following [110]	A robot takes a video as input, then builds a 3D dense map based on SLAM, and estimates the gaze simultaneously	Attention prediction is done by attention heat map adding occlusion and salience detection	The robot verbally describes the region focused on by the child based on attention heat maps
Intention understanding based on visual-NLP semantics and responding to the child accordingly [109]	The robot extracts information from one of the three types (i.e., vague, feeling, and clear) of instructions from a child and identifies the objects in front of the robot	The robot infers the intention of the child by a similarity computation algorithm and then transforms it into a structured robot control language	Turning itself to face the child, the robot points out a target object with a verbal description of the intention of the child

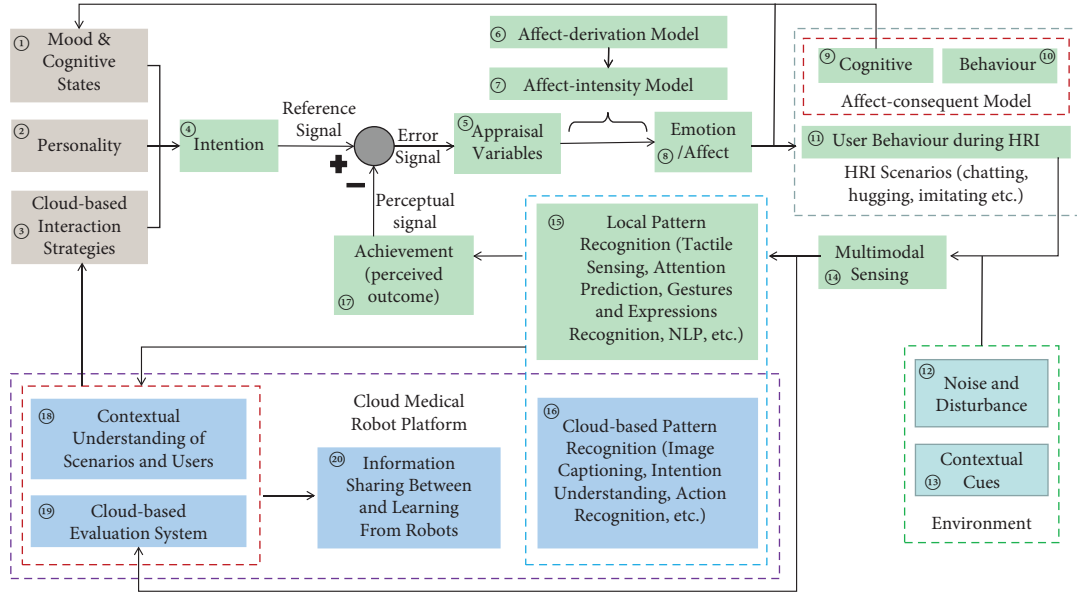


FIGURE 6: Compositional submodels (marked in red circled numeric symbol) of our proposed model AppraisalCloudPCT: (1) the submodels (i.e., number 4–17) marked in the green color theme constitute the main recursive control loop of a robot’s processing of emotional information, adopting perceptual control theory on emotions and a component model view of appraisal models; (2) the submodels (i.e., number 1–3) marked in the yellow-brown color theme constitute a robot’s intention of how to appraise an event (i.e., appraisal patterns of an interaction process), in which the appraisal patterns of the nine appraisal dimensions of a robot’s emotion can be affected by a robot’s mood and personality, and the interaction strategies; (3) the submodels (i.e., number 16–20) marked in the green color theme constitute the “cloud medical robot platform”; (4) the submodels (i.e., number 12–13) marked in the cyan color theme indicate that a robot will not only take the impact of the external environment (such as noise, disturbance, contextual cues) on the user during the interaction into account but also respond to the surrounding contexts appropriately.

through the linear regression analysis in [120]. And three equations of temperament including pleasure, arousal, and dominance are summarized in [122] as follows:

$$\begin{aligned} P\alpha &= 0.21E + 0.59A + 0.19N, \\ P\beta &= 0.15O + 0.30A - 0.57N, \\ P\gamma &= 0.25O + 0.17C + 0.60E - 0.32A, \end{aligned} \quad (5)$$

where $P\alpha$ denotes the value for the pleasant axis (α -axis), $P\beta$ denotes the value for the arousal axis (β -axis), and $P\gamma$ denotes the value for the dominance axis (γ -axis), respectively. Furthermore, the five factors of personality, i.e., O , C , E , A , $N \in [-1, 1]$, where O for openness, C for conscientiousness, E

for extraversion, A for agreeableness, and N for neuroticism, respectively.

The relationships between the five factors of personality and the appraisal dimensions of emotion could be derived in [121] (Page 519), where 10 main appraisal dimensions in major appraisal theories (Pleasantness, Goal Conduciveness, Effort, Perceived Control, Certainty, Agency-Self, Agency-Others, Agency-Circumstances, Unfairness, and Moral Violation), plus a new appraisal, relationship-involvement, were selected (see the Appendix in [121] for more details). Similarly, 9 personality-appraisal relationships (no relationship was found for appraisals “effort” and “relationship-involvement”) in [121] (Page 519) can be summarized as follows:

$$\text{Pleasantness}(F_{pl}) = -0.585N + 0.606C, \quad (6)$$

$$\text{Goal - Condu civeness}(F_{gc}) = -0.579N + 0.369C, \quad (7)$$

$$\text{Perceived Control}(F_{pc}) = -1.281N + 0.923E + 1.306C, \quad (8)$$

$$\text{Certainty}(F_c) = -1.203N + 0.880C, \quad (9)$$

$$\text{Agency - Self}(F_{as}) = -0.808A, \quad (10)$$

$$\text{Agency - Others}(F_{ao}) = -0.965C + 0.950O, \quad (11)$$

$$\text{Agency} - \text{Circumstances}(F_{ac}) = -0.587C, \quad (12)$$

$$\text{Unfairness}(F_u) = 1.149N - 0.928E - 1.113C, \quad (13)$$

$$\text{Moral Violation}(F_{mv}) = 1.309N - 1.005E - 1.456C - 0.840O, \quad (14)$$

where $O, C, E, A, N \in [-1, 1]$

Each equation indicates a relationship between an appraisal dimension and a combination of the Big Five personality traits, i.e., the tendency of appraising events in the particular appraisal dimension by people with specific personality traits. For instance, in equations (6) and (7), people with low N and high C will be more likely to appraise events as pleasant (Pleasantness) and as conducive to important goals (Goal-Conduciveness), although the tendency of appraising the same event in the two appraisal dimensions is not exactly the same. Note that once the value of the Big Five personality traits is determined, the value of each appraisal dimension will be also determined in equations (6)–(14).

5.2.3. The 3rd Submodel “Cloud-Based Interaction Strategies”.

The main purpose of this submodel is to output a strategy that a robot can use in the next round of interaction with an autistic child. Adopting the perceptual control theory on emotions, our proposed model AppraisalCloudPCT is designed in the first place to enable many rounds of recursive interaction between a robot and an autistic child, so that the interaction will be more effective, efficient, and easier to be satisfied by the child. By “strategy,” it means that, given the specific estimation of valence, arousal, and engagement levels of the child supported by the submodel “cloud-based evaluation system” and the contextual understanding of the interactive scenario and the child supported by the submodel “contextual understanding of scenarios and users,” the robot will be able to alter its mood and personality to match with the status of the child and the interactive context, for a better round of interaction.

As mentioned above in 5.2.2, for a better performance in human-robot interaction, a robot should have a similar personality to human participants, and the effects of context should be taken into consideration when designing a robotic personality. In this study, it is, therefore, necessary for the robot to have knowledge of the personality profile of an autistic child (this can be supported by the 19th submodel “cloud-based evaluation system,” as illustrated in Figure 7 that personality profile can be provided by the child’s parents) and to understand the interactive scenario and the child in the surrounding context (this can be supported by the 18th submodel “contextual understanding of scenarios and users,” please refer to it for more details).

Consequently, this submodel will output cloud-based interaction strategies as follows:

Strategy one: To match a robot’s personality with that of a child, first, the personality profile (i.e., rating scales of O, C, E, A, N between -1 and 1) of an autistic child who will interact with the robot will be obtained, and then, a robot’s personality will match with the child’s personality. Once the personality of the robot is altered, the emotional tendency that the robot will be experiencing and the appraisal patterns that the robot will use can be predicted by using the (10) equations in 5.2.1 and 5.2.2.

Strategy two: As contexts effect of a user’s perception of not only the emotions but also the personality of a robot, effects of context should be taken into consideration. First, the role that the robot plays in the task of the HRI scenario and what kind of personality that an autistic child expects to be consistent with such a task or role should be identified in the first place. Then, the personality of the robot should be modified to adapt to the child’s expectation.

Strategy three: The outcome of “contextual understanding of scenarios and users” should be taken into account, given that the noise and disturbance, and contextual cues may influence an autistic child’s mood and his/her judgement of the robot’s emotions. To do that, first, the emotional valence of the contextual cues will be obtained. Then, the robot’s mood should be congruent with the emotional valence of the contextual cues to some extent. Thirdly, in case of noise and disturbance were detected in the HRI scenario, the robot’s estimation of the child’s valence and arousal levels provided by the submodel “cloud-based evaluation system” should be rectified to some extent depending on the amount of the noise and disturbance.

5.2.4. The 4th Submodel “Intention”. “Intention” in this submodel means how will a robot intends to appraise an event (i.e., appraisal patterns of an interaction process), based on a robot’s mood and personality with the consideration of an interaction strategy for the next round of interaction. The main purpose of this submodel is to map the outputs of the first three submodels, namely, “mood and cognitive states,” “personality,” “cloud-based interaction strategies,” into appraisal patterns, which can be defined as follows:

$$F_{\text{intention}} = (F_{pl} + \Delta pl, F_{gc} + \Delta gc, F_{pc} + \Delta pc, Fc + \Delta c, F_{as} + \Delta as, F_{ao} + \Delta ao, F_{ac} + \Delta ac, Fu + \Delta u, F_{mv} + \Delta mv), \quad (15)$$

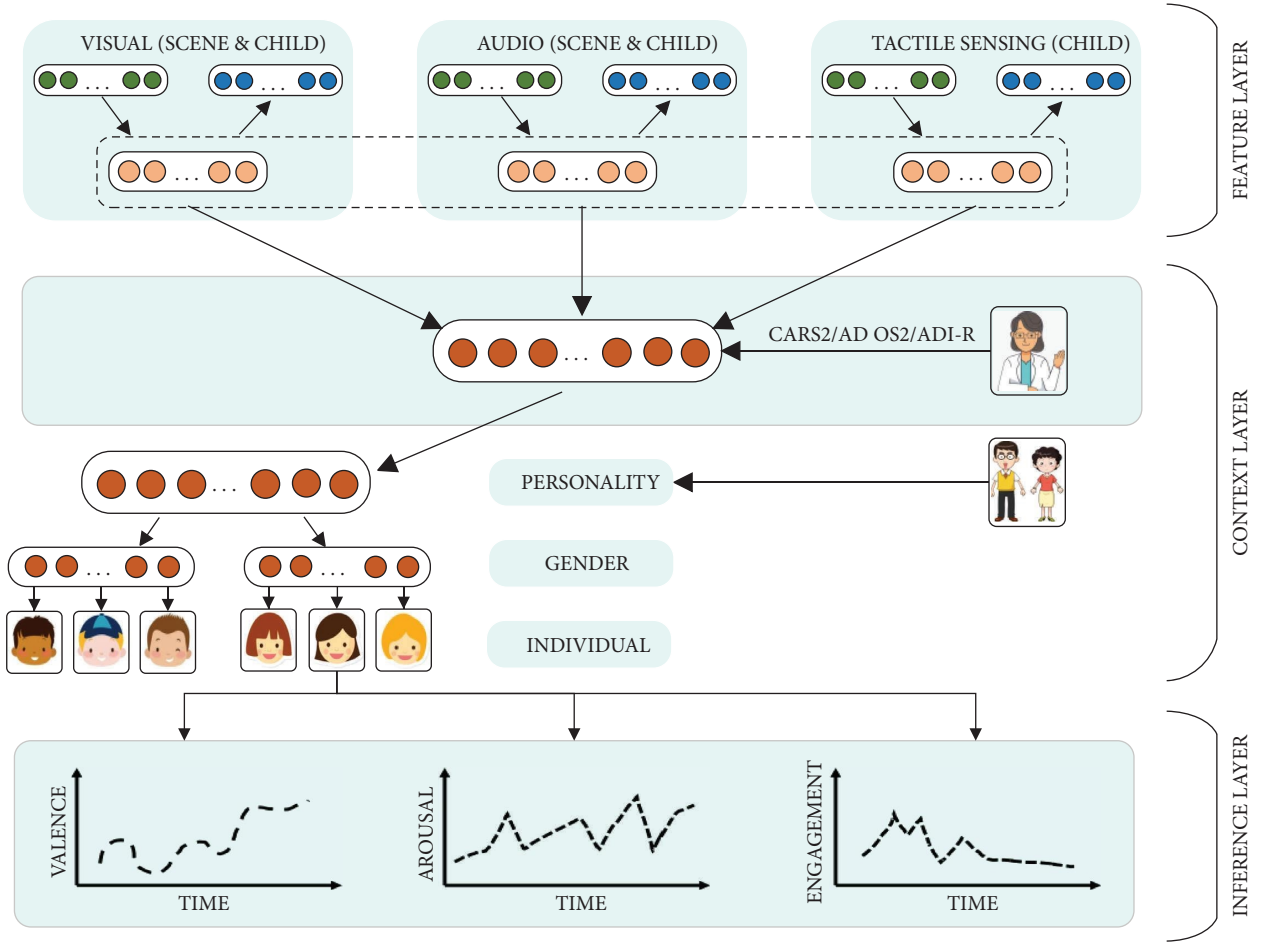


FIGURE 7: Three layers of modified PPA-net based on the work in [123]: (1) feature fusion is performed in the feature layer using features from three modalities (visual, audio, and tactile); (2) the context layer firstly uses behavioral scores of the child’s verbal ability, motor, and mental, to augment the input features using the autistic rating scales such as CARS2 [124], ADOS-2 [125], and ADI-R [126], and then, the GPA-NET (group-level network) is trained and used to initialize the personalized PPA-net weights at the personality, gender, and individual level (using clone); (3) the third layer is the inference layer, in which the child-specific estimation of valence, arousal, and engagement levels will be performed.

where F_{pl} , F_{gc} , F_{pc} , F_c , F_{as} , F_{ao} , F_{ac} , F_u , F_{mv} represent Pleasantness, Goal Conduciveness, Perceived Control, Certainty, Agency-Self, Agency-Others, Agency-Circumstances, Unfairness, and Moral Violation, respectively, as defined in Equation (6)–(14) in 5.2.2, and Δpl , Δgc , Δpc , Δc , Δas , Δao , Δac , Δu , Δmv represent the impact of the two submodels “mood and cognitive states” and “cloud-based interaction strategies” on the tendency of appraising events in the particular appraisal dimension, respectively.

5.2.5. The 5th Submodel “Appraisal Variables”. As mentioned in 4.1.2, in a closed-loop as illustrated in Figure 1, a collection of intention of a robot (i.e., a reference signal) and achievement (perceived outcome) (i.e., a perceptual signal) will cause an experienced emotion, and at the same time, an output-caused change in the cognitive states and behavior of the robot will affect a user’s behavior during the human-robot interaction. In other words, the difference (i.e., a mismatch) between the reference signal and the perceptual

signal will immediately result in an error signal, which will give rise both to emotional behavior and thinking of a robot.

Appraisal variables are defined as the set of specific judgments by which a robot can generate different emotional responses. The main purpose of this submodel is to output the error signal (i.e., a mismatch between a collection of intention of the robot and the achievement (perceived outcome)) as appraisal variables. Here, the error signal can be defined as follows:

$$F_{\text{error}} = F_{\text{intention}} - F_{\text{perceived}}, \quad (16)$$

where $F_{\text{intention}}$ represents a collection of intention of the robot as defined in equation (15) in 5.2.4, and $F_{\text{perceived}}$ represents the achievement (perceived outcome) of the robot as defined in equation (18) in 5.2.17.

5.2.6. The 6th Submodel “Affect-Derivation Model”. This submodel will specify the mapping from appraisal variables to affective state, and once a pattern of appraisals has been

determined how a robot will react emotionally. According to Itoh et al. [112], emotion vector $E = (E_p, E_a, E_d)$ can be expanded into the second-order differential equation as shown in equation (17) as follows:

$$M\ddot{E} + \Gamma\dot{E} + KE = F_{EA}, \quad (17)$$

where M , Γ , K , F_{EA} represent the emotional inertia matrix, emotional viscosity matrix, emotional elasticity matrix, and emotional appraisal, respectively. And the emotional appraisal FEA stands for the total result of appraising the appraisal variables (i.e., the error signal Error). According to Itoh et al. [112], by changing the emotional coefficient matrixes, the robot can express different reactions to a same stimulus.

5.2.7. The 7th Submodel “Affect-Intensity Model”. The strength of the emotional response resulting from a specific appraisal is specified in this submodel. As mentioned in 4.1.2, emotions with greater intensity and longer duration will arise in connection with a larger error that demands a robot to alter its affect-consequent model more appropriately to correct the error. Therefore, the bigger the error signal Error becomes, the greater the intensity of and with longer duration an emotion will be, and the stronger the emotional response will be.

5.2.8. The 8th Submodel “Emotion/Affect”. For each discrete emotion the robot will be experiencing, emotion vector $E = (E_p, E_a, E_d)$ can be mapped in the PAD mental space consisting of the pleasantness, arousal, and dominance axes. For mood vector, $M = (M_p, M_a, 0)$ consisting of a pleasantness component M_p and an arousal component M_a can also be mapped in the PAD mental space.

5.2.9. The 9th Submodel “Cognitive-Consequent Model”. This submodel determines how affect alters the nature or content of cognitive processes such as a robot’s beliefs, desires, and intentions, respectively. As mentioned above, an error signal (i.e., a mismatch between the intention of the robot and the achievement (perceived outcome)) will result in a robot’s intention to correct the error. How strong will the intention to correct the error be depends on how big the error is. Furthermore, as the robot is experiencing an emotion, its mood will be effected to some extent.

5.2.10. The 10th Submodel “Behavior-Consequent Model”. This submodel summarizes how affect alters our robot’s observable physical behavior such as facial expressions. As described in Table 2 in Section 5.1, our robot is equipped with 6 key capabilities in interactive scenarios with Chinese autistic children, and it can express itself to the children through facial expressions, gestures, speech, etc. In the interactive scenarios, the robot should alter its observable physical behavior in a manner, according to not only the emotion it is experiencing but also the three cloud-based interaction strategies described in 5.2.3.

5.2.11. The 11th Submodel “User Behavior during HRI”. In the child-robot interaction scenarios (e.g., having a conversation, hugging, playing games), an autistic child will generate certain behavior to adopt to/finish/withdraw from the child-robot interaction. Such behavior (e.g., gaze regulation, facial expressions, hand and body gestures, verbal expression), not only is a product of the child-robot interaction but also can be effected by the outward behavior of the robot as defined in the submodel “behavior-consequent model.”

5.2.12. The 12th Submodel “Noise and Disturbance”. Noise in this submodel is defined as noise coming from the surrounding contexts (e.g., ambient noise, other human voices other than the voice of the autistic child during the child-robot conversation). And disturbance is defined as any unexpected event that will have an adverse impact on the child-robot conversation, such as a heavy push on the robot, and the autistic child is forced by somebody to end the child-robot interaction in advance. Both the noise and disturbance can be detected by the sensors to perceive the child and the environment and by the self-checking sensors (e.g., torque sensors) inside the robot.

5.2.13. The 13th Submodel “Contextual Cues”. Robot faces can be viewed in the same way as human faces, according to [102] that users’ perceptions of a robot’s simulated emotional expressions can be affected by different emotional surrounding contexts (i.e., consistent or inconsistent classical music, or BBC news). Furthermore, when there is emotional context around, people are more able to recognize the facial expressions of the robot when the emotional valence of the environment is consistent with the facial expressions of the robot than when the emotional valence of the environment and its facial expressions are not consistent [99–101].

Consequently, it is important for the robot to perceive the emotional valence (i.e., contextual cues) of the surrounding contexts (e.g., sound, music, pictures/posters on the wall, video clips on the TV) in the interactive scenarios. As such, contextual cues will be considered, collected, and added to our proposed model AppraisalCloudPCT in this submodel.

5.2.14. The 14th Submodel “Multimodal Sensing”. In this submodel, the robot will perceive the autistic child and sense the environment through various sensors (e.g., camera, microphone arrays, tactile sensing arrays, infrared sensor) and multiple modalities (e.g., visual, auditory, and tactile sensing). A collection of sensor data in this submodel will feed to two submodels “local pattern recognition” and “cloud-based pattern recognition” and will be uploaded to the cloud medical robot platform, more specifically, to the submodel “cloud-based evaluation system.”

5.2.15. The 15th Submodel “Local Pattern Recognition”. In this submodel, our proposed model AppraisalCloudPCT will

output the results of the local pattern recognition (i.e., processes that run on the NVIDIA Jetson TX2 inside the robot body, as illustrated in Figure 3), and the results will be uploaded to the cloud medical robot platform to facilitate the two submodels, i.e., the cloud-based evaluation system and the contextual understanding of scenarios and users.

The output of tactile sensing is defined as $TS = (P_i, TB_j)$, where P_i is the i th position of the robot body being touched by an autistic child $P_i \in \{\text{Top of Head, Back of Head, Forehead, Left Cheek, Right Cheek, Front of Right Forearm, Back of Right Forearm, Front of Right Upper Arm, Back of Right Upper Arm, Back of Left Upper Arm, Front of Left Upper Arm, Back of Left Forearm, Front of Left Forearm, Right Rear Hip, Back of Right Thigh, Inner Right Thigh, Lower Right Thigh, Left Rear Hip, Back of Left Thigh, Inner Left Thigh, Lower Left Thigh}\}$, and TB_j is the j th touching behavior pattern of the autistic child $TB_j \in \{\text{Palm Momentary Sliding, Palm Momentary Tapping, Random Finger Poking, Finger Sliding, Random Slow Sliding, Random Momentary Tapping}\}$.

The output of Attention Prediction (i.e., gaze and head direction estimation) is defined as $AP = (dl, dr, dh)$, where dl and dr are gaze direction of the left and right eyes of an autistic child respectively, and parameter dh represents the head direction.

The output of gestures recognition is defined as $GR = (HG_i, BG_j)$, where HG_i is the hand gesture of an autistic child $HG_i \in \{\text{OK, Peace, Punch, Stop, Nothing}\}$, and BG_j is the body gesture of the autistic child $BG_j \in \{\text{Standing, Walking, Running, Jumping, Sitting, Squatting, Kicking, Punching, Waving, None}\}$.

The output of facial expressions recognition is defined as $FE_i \in \{\text{Happiness, Sadness, Anger, Surprise, Fear, Disgust, Neutral}\}$.

The output of natural language processing (NLP) is not defined as the whole sentences in a conversation between the robot and an autistic child, but as pertinent words or word stems in natural languages that can commonly distinguish 36 affective categories, as defined in [57] (Page 714–715). Therefore, $NLP_{\text{output}} = (PWs, AC)$, where PWs represents all of the pertinent words or word stems as defined in [57] that can be extracted from a conversation and affective category $AC \in \{\text{Contentment, Anger, Admiration/Awe, Anxiety, Amusement, Being, Touched, Desperation, Boredom, Compassion, Contempt, Disappointment, Disgust, Dissatisfaction, Envy, Fear, Feeling, Gratitude, Guilt, Happiness, Hatred, Hope, Humility, Interest/Enthusiasm, Irritation, Jealousy, Joy, Longing, Lust, Pleasure/Enjoyment, Pride, Relief, Sadness, Relaxation/Serenity, Tension/Stress, Shame, Surprise, Positive, Negative, Neutral}\}$ (36 affective categories plus Neutral).

5.2.16. The 16th Submodel “Cloud-Based Pattern Recognition”. In this submodel, our proposed model AppraisalCloudPCT will output the results of the cloud-based pattern recognition (i.e., processes that run on the cloud, as illustrated in Figure 3), and the results will be uploaded to the cloud medical robot platform to facilitate the

three submodels, i.e., the cloud-based evaluation system, the contextual understanding of scenarios and users, and the information sharing between and learning from robots.

The output of image captioning is defined as $IC = (Ob, Pr, At)$, where Ob represents the object concentrated by an autistic child, the region of which can be represented by an attention heat map of an image captured by the robot camera, Pr represents the preposition, and At represents the attributes of the object.

The output of intention understanding is defined as $IU = (Insi, TO, DP, RCL)$, where $Insi$ is one of the three types of natural language instructions given by an autistic child $Insi \in \{\text{Clear Type, Vague Type, Feeling Type}\}$, TO represents the target object out of multiple objects in front of the robot, DP represents the delivery place that the target object should be delivered to, and the RCL format utilized in this paper is “Grasp TO to DP,” which is the structured language that can be comprehended by robots.

The output of action recognition is defined as $AR = (AB_i, HB_j)$, where AB_i belongs to 10 kinds of abnormal behaviors of an autistic child plus the normal status $AB_i \in \{\text{Clapping Hands, Swinging Back and Forth, Spinning Circles, Flipping Fingers, Bumping Heads, Clapping Ears, Turning Fingers, Scratching, Walking on Tiptoe, Snapping Fingers, Normal Status}\}$ and HB_j belongs to 5 kinds of unhealthy conditions plus the healthy status $HB_j \in \{\text{Falling Down, Having Headache, Having Chest and Abdominal Pain, Having Back Pain, Having Neck Pain, Healthy Status}\}$.

5.2.17. The 17th Submodel “Achievement (Perceived Outcome)”. The importance of this submodel is to summarize the feedback (e.g., the interpretation of an autistic child’s intention, attention, emotional state, and behavior) provided by the child during/after the human-robot interaction. The achievement (perceived outcome) of the robot can be defined as follows:

$$F_{\text{perceived}} = (PF_{pl}, PF_{gc}, PF_{pc}, PF_c, PF_{as}, PF_{ao}, PF_{ac}, PF_u, PF_{mv}), \quad (18)$$

where PF_{pl} , PF_{gc} , PF_{pc} , PF_c , PF_{as} , PF_{ao} , PF_{ac} , PF_u , PF_{mv} represent the 9 appraisal dimensions respectively as described in 5.2.2, i.e., Pleasantness, Goal Conduciveness, Perceived Control, Certainty, Agency-Self, Agency-Others, Agency-Circumstances, Unfairness, and Moral Violation that will be used to appraise the achievement (perceived outcome). These 9 appraisal dimensions will be defined in equation (19)–with PF_{pl} as follows:

$$PF_{pl} = a_1 \cdot OT_1 + a_2 \cdot OT_2 + a_3 \cdot OT_3 + a_4 \cdot OT_4 + a_5 \cdot OT_5, \quad (19)$$

where $OT_1, OT_2, OT_3, OT_4 \in [-1, 1]$, and $OT_5 \in [-1, 0]$ represent OutcomeType1, OutcomeType2, OutcomeType3, OutcomeType4, OutcomeType5, respectively, and a_1, a_2, a_3, a_4, a_5 are the coefficient of each outcome type.

In this submodel, we categorize the achievement (perceived outcome) into 5 types: ① OutcomeType1: “Friendly VS. Unfriendly” type, e.g., “Friendly” in the outcome of

“Tactile Sensing” and “Gestures Recognition” means that the interpretation of the attitude of an autistic child towards the robot would be friendly, and an extreme friendly outcome, a neutral outcome, and an extreme friendly outcome of this type will be 1, 0, and -1 , respectively; ② OutcomeType2: “Positive VS. Negative” type, e.g., “Positive” in the outcome of “Facial Expressions Recognition,” “Natural Language Processing,” and “Contextual Cues” means that, the emotional valence would be positive (e.g, output of a “dislike” in “Natural Language Processing” will be categorized as “Negative”), and an extreme positive outcome, a neutral outcome, and an extreme negative outcome of this type will be 1, 0, and -1 respectively; ③ OutcomeType3: “Valid VS. Invalid” type, e.g., “Valid” in the outcome of “Image Captioning” and “Intention Understanding” means that an autistic child will react positively after the robot verbally described the objects in the image or the robot verbally stated the intention in the interactive scenarios, and an extreme valid outcome, a no feedback outcome, and an extreme invalid outcome of this type will be 1, 0, and -1 respectively; ④ OutcomeType4: “Focused VS. Distracted” type, e.g., “Focused” in the outcome of “Attention Prediction” means that, during the human-robot interaction, the robot can predict that an autistic child has “focused” on one or two objects in the interactive scenario; on the contrary, “Distracted” means the gaze and head direction of the

child cannot “fixed on” one or two objects, rather they shifted from one object to another object too often, and “None” means the child cannot “focused” on any object, and an extreme focused outcome, a none outcome, and an extreme distracted outcome of this type will be 1, 0, and -1 respectively; ⑤ OutcomeType5: “Normal VS. Unnormal” type, e.g., “Unnormal” in the outcome of “Action Recognition” and “Noise and Disturbance” means that the robot can detect some abnormal/unhealthy behavior (e.g., “walking on tiptoe” or “having back pain”) of the child or some noise/disturbance in the interactive scenarios, and a normal outcome, and an extreme unnormal outcome of this type will be 0 and -1 , respectively.

Note that the probability of simultaneous occurrence of most of or all of these types of outcomes is very low, and usually only a few of them will occur. For each kind of the pattern recognition (i.e., pattern recognition in submodels “local pattern recognition” and “cloud-based pattern recognition”) and the sensing of the environment (i.e., the sensing in submodels “noise and disturbance” and “contextual cues”), as described in the above submodels, the outcome value of which will be mapped into $[-1, 1]$ or $[-1, 0]$ using fuzzy sets depends on which type of outcome is categorized as follows.

Similarly, PF_{gc} , PF_{pc} , PF_c , PF_{as} , PF_{ao} , PF_{ac} , PF_u , PF_{mv} can be defined as follows:

$$PF_{gc} = a_1 \cdot OT_1 + a_2 \cdot OT_2 + a_3 \cdot OT_3 + a_4 \cdot OT_4 + a_5 \cdot OT_5, \quad (20)$$

$$PF_{pc} = b_1 \cdot OT_1 + b_2 \cdot OT_2 + b_3 \cdot OT_3 + b_4 \cdot OT_4 + b_5 \cdot OT_5, \quad (21)$$

$$PF_c = c_1 \cdot OT_1 + c_2 \cdot OT_2 + c_3 \cdot OT_3 + c_4 \cdot OT_4 + c_5 \cdot OT_5, \quad (22)$$

$$PF_{as} = d_1 \cdot OT_1 + d_2 \cdot OT_2 + d_3 \cdot OT_3 + d_4 \cdot OT_4 + d_5 \cdot OT_5, \quad (23)$$

$$PF_{ao} = e_1 \cdot OT_1 + e_2 \cdot OT_2 + e_3 \cdot OT_3 + e_4 \cdot OT_4 + e_5 \cdot OT_5, \quad (24)$$

$$PF_{ac} = f_1 \cdot OT_1 + f_2 \cdot OT_2 + f_3 \cdot OT_3 + f_4 \cdot OT_4 + f_5 \cdot OT_5, \quad (25)$$

$$PF_u = g_1 \cdot OT_1 + g_2 \cdot OT_2 + g_3 \cdot OT_3 + g_4 \cdot OT_4 + g_5 \cdot OT_5, \quad (26)$$

$$PF_{mv} = h_1 \cdot OT_1 + h_2 \cdot OT_2 + h_3 \cdot OT_3 + h_4 \cdot OT_4 + h_5 \cdot OT_5. \quad (27)$$

5.2.18. *The 18th Submodel “Contextual Understanding of Scenarios and Users”*. As illustrated in Figure 6, the outcome of each kind of the pattern recognition and the sensing of environment will be uploaded to the “cloud medical robot platform,” more specifically, to this submodel and the next submodel “cloud-based evaluation system.” As such, in this submodel, our proposed model AppraisalCloudPCT will output the outcome of “contextual understanding of scenarios and users,” which is defined as $CUSU = (US, UU)$, where US represents the understanding of scenarios provided mainly by the output of image captioning and of sensing the environment (i.e., combing scene description

with the sensing of noise and disturbance, and of contextual cues), and UU represents the understanding of users provided mainly by the output of other local and cloud-based pattern recognition (i.e., gaze estimation, intention, gestures).

5.2.19. *The 19th Submodel “Cloud-Based Evaluation System”*. The importance of this submodel is to provide insights into both the cognitive and behavioral status of an autistic child, and of the intention of the child to engage with the robot, to the submodel “cloud-based interaction strategies.”

In this submodel, a personalized machine learning (ML) framework, so-called the personalized perception of affect network (PPA-net) developed by an MIT research group [123], will be adopted in the “cloud-based evaluation system.” As illustrated in Figure 7, in the modified PPA-net, group-level perception of affect network (GPA-net) is trained with the data exacted from the autistic rating scales provided by the doctor or therapist of the child, and the data exacted from the personality profile of the child provided by the parents of the child. Consequently, by using the modified PPA-net, this submodel can automatically provide a continuous and simultaneous estimation of levels of engagement and affective states (i.e., arousal and valence) of an autistic child, to the submodel “cloud-based interaction strategies.”

5.2.20. The 20th Submodel “Information Sharing Between and Learning from Robots”. As mentioned earlier in chapter 4.1.2, one advantage in the research on cloud medical robots is data of interaction between a user and a robot can be stored and evaluated in a cloud for further assessment of the social and communicative characteristics of a user. With the cloud medical robot platform, in this submodel, information (e.g., the personality profile of each autistic child) can be shared between robots with the support of the submodel “cloud-based evaluation system,” and the capability of interpretation of a user and of making decisions can be learned with the support of the submodel “contextual understanding of scenarios and users.”

6. Conclusions, Discussion, and Future Work

6.1. Conclusions. In this article, we present a novel computational model of emotions so-called AppraisalCloudPCT for socially interactive robots, especially for robots for a special group of people such as autistic children. This model takes into account the social and communicative characteristics of autistic children so that it can fit the need of the autistic children. It mainly results from that our proposed model not only has solid theoretical ground built on a component view of computational models, the appraisal theories on emotions, cloud robotics, and perceptual control theory on emotions but also can be implemented in a social robot developed by us for autistic rehabilitation by adopting mood equation, emotion equation, and personality equation.

Moreover, compared to other significant computational models of emotions for socially interactive robots, our proposed model AppraisalCloudPCT has a number of merits. First, our proposed model can guarantee sufficient rounds of recursive interaction between a robot and an autistic child, so that the interaction will be more effective, efficient, and easier to be satisfied by the child. Second, with our proposed model, a robot can simulate the whole process of human emotion (e.g., generation, regulation, and responding to a stimulus) to a great extent. Third, our proposed model can facilitate sharing information between and learning from various socially interactive robots. Last

but not least, our proposed model can be highly computable so that it is suitable to be implemented in various socially interactive robots.

6.2. Limitations. Our proposed model AppraisalCloudPCT is designed based on Marsella et al.’s compositional view of model building [5], which lays stress on that emotional models are often composed of individual “submodels” or “smaller components” that can be matched, mixed, or excluded from any given implementation and are often shared. According to Marsella et al. [5], components may be evaluated and subsequently abandoned or improved due to ongoing evaluations before the final version of the model is designed. Although our model is completely designed, there is still room for finding alternative or better mathematical definitions, equations, or algorithms for realizing each individual “submodels.”

6.3. Discussion. In this article, we proposed a novel computational model of emotions called AppraisalCloudPCT and elaborated on how to implement it in a socially interactive robot we developed for autistic rehabilitation. However, there are several points that are worthy of being addressed as follows.

First of all, this study is aimed specifically at designing the computational model of emotions for autistic children-robot interaction for three reasons as follows. (1) Although minimal progress has been made in advancing the clinical use of robotics in ASD interventions in clinical settings [13], applying robots for autism interventions still achieved a number of targets [11], and 24 of 74 ASD objectives in the “eight domains” as mentioned in Section 1 can potentially be applied to. (2) Modeling of emotions is of critical importance for robots when interacting socially with humans [8]. This is so because the robot’s emotional responses are determined by the robot’s computational model of emotion, in the light of its own internal cognitive-affective state and its interactions with the external environment [7]. (3) There are four world-leading research groups with pioneering work in promoting social robots as useful tools in autism therapy, but none of them have designed or applied computational models of emotions for the social robots used in their autism therapy studies.

Second, in Section 4.2, we chose the five crucial properties of a computational emotion model as listed in [103], alongside with one more property, i.e., combining with cloud robotics, to be the six criteria for comparison. We believe that “combining with cloud robotics” can be a crucial property of a computational emotion model, and it can be a fair criterion for comparison of computational emotion models to make robots smarter and better satisfied by the users, as well as to promote sales in the service robots market for three reasons as follows. (1) As mentioned before, a computational model of emotions should endow a robot with a more powerful capability of making decisions faster, more appropriate, and more efficient, given that more and more socially interactive robots will be exposed to various users with different backgrounds and be connected to

substantial Internet of Things (IoT) such as medical IoT with massive medical data. As “combining with cloud robotics” can facilitate sharing information between and learning from socially interactive robots, we believe that this property will be crucial in building the computational models. (2) Given that other crucial properties such as (iv) data-driven mapping and (v) ethical reasoning are heavily data-driven and in great demand of computing power, “combining with cloud robotics” could be an efficient if not the best way to guarantee that data consisting of interaction between a user and a robot can be stored and evaluated in a cloud for further assessment of the social and communicative characteristics of a user. (3) On one hand, more and more socially interactive robots are implemented artificial intelligence (AI) algorithms or deep learning (DL) (e.g., the modified PPA-net implemented in our own robot)/reinforcement learning (RL)/deep reinforcement learning (DRL) frameworks to make them smarter and better received by the users; on the other hand, deploying them in the main controller of a robot rather than in a cloud will increase the hardware cost due to increased computational load. Since the parents of autistic children usually suffer from heavy burden not only mentally but financially, “combining with cloud robotics” would be necessary for promoting robots with acceptable prices in the service robots market to those parents.

Third, our proposed model AppraisalCloudPCT could be implemented in a socially interactive robot that we developed for autistic rehabilitation. Such a model could also be adapted to service people with different special needs, e.g., dementia elders. This results from that our proposed computational model of emotions takes into account the social and communicative characteristics of a special group of users such as autistic children or dementia elders, through its submodel so-called cloud-based interaction strategies, which is supported by two submodels (i.e., “contextual understanding of scenarios and users” and “cloud-based evaluation system”) in a cloud medical robot platform, as illustrated in Figure 1. As mentioned before, a cloud-based evaluation system enables the data of interaction between a user and a robot to be stored and evaluated in a cloud for further assessment of the social and communicative characteristics of a user. Furthermore, the submodel “contextual understanding of scenarios and users” relies on another two submodels “local pattern recognition” and “cloud-based pattern recognition,” which can provide the interpretation of a user’s intention, attention, emotional state, and behavior. Therefore, the proposed computational model of emotions is suitable for socially interactive robots, particularly robots for a special group of users such as autistic children or dementia elders, which promote the universality of our model to some extent. Moreover, our proposed computational model also meets the criteria of “domain-independent,” i.e., processing and exhibiting emotional responses in various situations as well as in certain kinds of interaction domain, since it can coordinate a robot implemented with our model to respond to the surrounding emotional contexts appropriately.

For our proposed model to be adapted to socially interactive robots servicing dementia elders, a few steps would

be necessary as follows. (1) As illustrated in Figure 7, a group-level perception of affect network (GPA-net) in the modified PPA-net will be trained with the data exacted from the dementia rating scales such as mini-mental state examination (MMSE) [127] provided by the doctor or therapist of the dementia elder, and the data exacted from the personality profile of the dementia elder provided by the offspring or close friends of the dementia elder. Consequently, by using the modified PPA-net, this submodel can automatically provide simultaneous and continuous estimation of the different levels of affective states (i.e., valence and arousal) and engagement of a dementia elder, to the submodel “cloud-based interaction strategies.” (2) With the support from the two submodels “cloud-based evaluation system” and “contextual understanding of scenarios and users” which can provide the specific estimation of valence, arousal, and engagement levels of the dementia elder, and the contextual understanding of the interactive scenario and the dementia elder, respectively, the robot will be able to alter its mood and personality to match with the status of the dementia elder and the interactive context using the three interaction strategies in the submodel “cloud-based interaction strategies,” for a better round of interaction. (3) Our proposed model is designed in the first place to enable many rounds of recursive interaction between a robot and a user. Based on the feedback (e.g., the interpretation of a dementia elder’s intention, attention, emotional state, and behavior) provided by the dementia elder during/after the human-robot interaction, as summarized by the submodel “achievement (perceived outcome)”, the three interaction strategies in the submodel “cloud-based interaction strategies” can be modified accordingly. As such, the interaction will be more effective, efficient, and easier to satisfy the needs of dementia elder after many rounds of recursive interaction.

6.4. Future Work. Future studies should examine how our model performs in various robots and in more interactive scenarios.

Data Availability

All data included in this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors thank Mr. Lifu Chen and his colleagues who are from the Smart Children Education Center (Shenzhen) and the DoGoodly International Education Center (Shenzhen) Co., Ltd., and Dr. Guobin Wan and his colleagues who are from Shenzhen Maternal and Child Health Hospital, for generously giving constructive suggestions in conducting standardized clinical studies with our developed socially interactive robot Dabao. This work was supported by the Shenzhen Science and Technology Innovation Commission

(Grants nos. JCYJ20180508152240368 and GJHZ20200731095205016).

References

- [1] J. Lin, S. Mare, and Z. Micheal, "Computational models of emotion and cognition," *Advances in Cognitive Systems*, vol. 59, 2012.
- [2] R. S. Lazarus, "Cognition and emotion in motivation," *American Psychologist*, vol. 46, no. 4, pp. 352–367, 1991.
- [3] H. A. Simon, *The Sciences of the Artificial*, The MIT Press, Cambridge, MA, USA, 2nd edition, 1981.
- [4] L. Canamero, "Embodied robot models for interdisciplinary emotion research," *IEEE Transactions on Affective Computing*, vol. 12, no. 2, pp. 340–351, 2021.
- [5] S. Marsella, J. Gratch, and P. Petta, "Computational models of emotion," *A blueprint for affective computing-A sourcebook and manual*, vol. 11, no. 1, pp. 21–45, 2010.
- [6] C. Breazeal, "Function meets style: insights from emotion theory applied to HRI," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 2, pp. 187–194, 2004.
- [7] C. Breazeal, "Emotion and sociable humanoid robots," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 119–155, 2003.
- [8] C. Breazeal, A. Takanishi, and T. Kobayashi, *Social Robots that Interact with People*, Springer Handbook of Robotics, Berlin, Germany, 2008.
- [9] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, American Psychiatric Publishing, Virginia, VA, USA, 5th edition, 2013.
- [10] G. B. Wan, F. H. Deng, Z. J. Jiang et al., "Attention shifting during child-robot interaction: a preliminary clinical study for children with autism spectrum disorder," *Frontiers of Information Technology and Electronic Engineering*, vol. 20, no. 3, pp. 374–387, 2019.
- [11] P. Pennisi, A. Tonacci, G. Tartarisco et al., "Autism and social robotics: a systematic review," *Autism Research*, vol. 9, no. 2, pp. 165–183, 2016.
- [12] C. A. G. J. Huijnen, M. A. S. Lexis, R. Jansens, and L. P. de Witte, "Mapping robots to therapy and educational objectives for children with autism spectrum disorder," *Journal of Autism and Developmental Disorders*, vol. 46, no. 6, pp. 2100–2114, 2016.
- [13] M. Begum, R. W. Serna, and H. A. Yanco, "Are robots ready to deliver autism interventions? a comprehensive review," *International Journal of Social Robotics*, vol. 8, no. 2, pp. 157–181, 2016.
- [14] S. Baron-Cohen, *Mind Blindness: An Essay on Autism and Theory of Mind*, MIT Press/Bradford Books, Boston, MA, USA, 1995.
- [15] O. Golan, E. Ashwin, Y. Granader et al., "Enhancing emotion recognition in children with autism spectrum conditions: an intervention using animated vehicles with real emotional faces," *Journal of Autism and Developmental Disorders*, vol. 40, no. 3, pp. 269–279, 2009.
- [16] A. Tapus, A. Peca, A. Aly et al., "Exploratory study: children's with autism awareness of being imitated by nao robot," *Interaction Studies*, vol. 13, no. 3, pp. 315–347, 2012.
- [17] C. A. Pop, R. Simut, S. Pinteá et al., "Can the social robot Probo help children with autism to identify situation-based emotions? A series of single case experiments," *International Journal of Humanoid Robotics*, vol. 10, no. 3, Article ID 1350025, 2013.
- [18] B. Robins and K. Dautenhahn, "Tactile interactions with a humanoid robot: novel play scenario implementations with children with autism," *International Journal of Social Robotics*, vol. 6, no. 3, pp. 397–415, 2014.
- [19] J. Wainer, K. Dautenhahn, B. Robins, and F. Amirabdollahian, "A pilot study with a novel setup for collaborative play of the humanoid robot kaspar with children with autism," *International Journal of Social Robotics*, vol. 6, no. 1, pp. 45–65, 2014.
- [20] C. A. G. J. Huijnen, M. A. S. Lexis, R. Jansens, and L. P. de Witte, "How to implement robots in interventions for children with autism? a co-creation study involving people with autism, parents and professionals," *Journal of Autism and Developmental Disorders*, vol. 47, no. 10, pp. 3079–3096, 2017.
- [21] B. A. English, A. Coates, and A. Howard, "Recognition of Gestural Behaviors Expressed by Humanoid Robotic Platforms for Teaching Affect Recognition to Children with Autism - A Healthy Subjects Pilot Study," in *Proceedings of the International Conference on Social Robotics, LNCS*, pp. 567–576, Tsukuba, Japan, November 2017.
- [22] B. Lee, J. Xu, and A. Howard, "Does appearance matter? Validating engagement in therapy protocols with socially interactive humanoid robots," in *Proceedings of the IEEE symposium series on computational intelligence (SSCI)*, pp. 1–6, Honolulu, HI, USA, November 2017.
- [23] J. Greczek and M. Mataric, "Encouraging user autonomy through robot-mediated intervention," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, pp. 189–190, Portland, OR, USA, March 2015.
- [24] M. J. Matorić, "Socially assistive robotics: human augmentation versus automation," *Science Robotics*, vol. 2, no. 4, Article ID eaam5410, 2017.
- [25] C. Clabaugh, D. Becerra, E. Deng, G. Ragusa, and M. Matorić, "Month-long, In-home Case Study of a Socially Assistive Robot for Children with Autism Spectrum Disorder," in *Proceedings of the Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 87–88, Chicago, IL, USA, March 2018.
- [26] R. Simut, G. V. D. Perre, and C. Costescu, "Probogotchi: a novel edutainment device as a bridge for interaction between a child with asd and the typically developed sibling," *Journal of Evidence-Based Psychotherapies*, vol. 16, no. 1, pp. 91–112, 2016.
- [27] P. G. Esteban, P. Baxter, T. Belpaeme et al., "How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder," *Paladyn. Journal of Behavioral Robotics*, vol. 8, no. 1, pp. 18–38, 2017.
- [28] C. A. Smith and P. C. Ellsworth, "Patterns of cognitive appraisal in emotion," *Journal of Personality and Social Psychology*, vol. 48, no. 4, pp. 813–838, 1985.
- [29] R. Mauro, K. Sato, and J. Tucker, "The role of appraisal in human emotions: a cross-cultural study," *Journal of Personality and Social Psychology*, vol. 62, no. 2, pp. 301–317, 1992.
- [30] I. J. Roseman, A. A. Antoniou, and P. E. Jose, "Appraisal determinants of emotions: constructing a more accurate and comprehensive theory," *Cognition and Emotion*, vol. 10, no. 3, pp. 241–278, 1996.
- [31] K. R. Scherer, "The role of culture in emotion-antecedent appraisal," *Journal of Personality and Social Psychology*, vol. 73, no. 5, pp. 902–922, 1997.

- [32] W. T. Powers, "Living Control Systems III: The Fact of Control," 2008, http://www.livingcontrolsystems.com/lcs3/content_lcs3.html.
- [33] J. Kuffner, "Cloud-Enabled Humanoid Robots," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robotics*, Nashville, TN, USA, December 2010.
- [34] B. Kehoe, A. Matsukawa, S. Candido, J. Kuffner, and K. Goldberg, "Cloud-based robot grasping with the google object recognition engine," in *Proceedings of the 2013 IEEE International Conference on Robotics and Automation*, pp. 4263–4270, Karlsruhe, Germany, May 2013.
- [35] M. Bonaccorsi, L. Fiorini, F. Cavallo, A. Saffiotti, and P. Dario, "A cloud robotics solution to improve social assistive robots for active and healthy aging," *International Journal of Social Robotics*, vol. 8, no. 3, pp. 393–408, 2016.
- [36] Cloud Standards Customer Council, "Impact of Cloud Computing on Healthcare Version 2.0," 2017, <https://www.omg.org/cloud/deliverables/CSCC-Impact-of-Cloud-Computing-on-Healthcare.pdf>.
- [37] E. Fosch-Villaronga, H. Felzmann, M. Ramos-Montero, and T. Mahler, "Cloud Services for Robotic Nurses? Assessing Legal and Ethical Issues in the Use of Cloud Services for Healthcare Robots," in *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 290–296, Madrid, Spain, October 2018.
- [38] Z. H. Khan, A. Siddique, and C. W. Lee, "Robotics utilization for healthcare digitization in global COVID-19 management," *International Journal of Environmental Research and Public Health*, vol. 17, no. 11, p. 3819, 2020.
- [39] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*, Cambridge University Press, Cambridge UK, 1990.
- [40] K. R. Scherer, "Appraisal considered as a process of multi-level sequential checking," *Appraisal processes in emotion: Theory, Methods, Research*, vol. 92, p. 120, 2001.
- [41] A. Mehrabian and J. A. Russell, *An Approach to Environmental Psychology*, The MIT Press, Cambridge, MA, USA, 1974.
- [42] A. Ortony, "On Making Believable Emotional Agents Believable," *Emotions in Humans and Artifacts*, pp. 189–211, The MIT Press, Cambridge, MA, USA, 2003.
- [43] C. Bartneck, "Integrating the OCC model of emotions in embodied characters," in *Proceedings of the Workshop on Virtual Conversational Characters: Applications, Methods, and Research Challenges*, Geneva, Switzerland, June 2002.
- [44] B. R. Steunebrink, M. Dastani, and J.-J. C. Meyer, "The OCC model revisited," in *Proceedings of the 4th Workshop on Emotion and Computing*, Memphis TN, USA, October 2009.
- [45] M. B. Arnold, "Emotion and Personality," 1960, <https://psycnet.apa.org/record/1960-35012-000>.
- [46] R. S. Lazarus, "Psychological Stress and the Coping Process," 1966, <https://psycnet.apa.org/record/1966-35050-000>.
- [47] K. R. Scherer, "Emotions are emergent processes: they require a dynamic computational architecture," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3459–3474, 2009.
- [48] K. R. Scherer, "On the Nature and Function of Emotion: A Component Process Approach," in *Approaches to Emotion*, K. R. Scherer and P. Ekman, Eds., pp. 293–317, Psychology Press, London, UK, 1984.
- [49] I. J. Roseman and C. A. Smith, "Appraisal theory," appraisal processes in emotion: theory, methods, research," pp. 3–19, 2001, <https://psycnet.apa.org/record/2001-06810-000>.
- [50] D. Sander, D. Grandjean, and K. R. Scherer, "A systems approach to appraisal mechanisms in emotion," *Neural Networks*, vol. 18, no. 4, pp. 317–352, 2005.
- [51] K. R. Scherer, "Appraisal Theories," *Handbook of Cognition and Emotion*, Wiley, Hoboken, NJ, USA, 1999.
- [52] P. M. Niedenthal, S. Kruth-Gruber, and F. Ric, "Theories of emotion," *The Psychology of Emotion: Interpersonal Experiential, and Cognitive Approaches, Principles of Social Psychology Series*, Psychology Press, New York, NY USA, 2006.
- [53] G. L. Clore and A. Ortony, "Cognition in emotion: always, sometimes, or never," *Cognitive Neuroscience of Emotion*, Oxford University Press, Oxford, UK, 2000.
- [54] K. R. Scherer, "Emotion as a multi-component process: a model and some cross-cultural data," *Emotions, relationships and health*, vol. 5, pp. 37–63, 1984.
- [55] K. R. Scherer, "Vocal affect expression: a review and a model for future research," *Psychological Bulletin*, vol. 99, no. 2, pp. 143–165, 1986.
- [56] K. R. Scherer, "Studying the emotion-antecedent appraisal process: an expert system approach," *Cognition and Emotion*, vol. 7, no. 3–4, pp. 325–355, 1993.
- [57] K. R. Scherer, "What are emotions? And how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2005.
- [58] J. A. Russell, "Core affect and the psychological construction of emotion," *Psychological Review*, vol. 110, no. 1, pp. 145–172, 2003.
- [59] L. F. Barrett, "Are emotions natural kinds?" *Perspectives on Psychological Science*, vol. 1, pp. 28–58, 2006.
- [60] D. Watson and A. Tellegen, "Toward a consensual structure of mood," *Psychological Bulletin*, vol. 98, no. 2, pp. 219–235, 1985.
- [61] W. James, "The Principles of Psychology," 1890, https://library.manipaldubai.com/DL/the_principles_of_psychology_vol_II.pdf.
- [62] S. Schachter and J. Singer, "Cognitive, social, and physiological determinants of emotional state," *Psychological Review*, vol. 69, no. 5, pp. 379–399, 1962.
- [63] J. A. Russell, "Emotion, core affect, and psychological construction," *Cognition and Emotion*, vol. 23, no. 7, pp. 1259–1283, 2009.
- [64] A. Scarantino, "Core affect and natural affective kinds," *Philosophy of Science*, vol. 76, no. 5, pp. 940–957, 2009.
- [65] C. Becker-Asano and I. Wachsmuth, "Affective computing with primary and secondary emotions in a virtual human," *Autonomous Agents and Multi-Agent Systems*, vol. 20, no. 1, pp. 32–49, 2010.
- [66] W. T. Powers and Behavior, *The Control of Perception*, Aldine de Gruyter, Chicago, CA, USA, 2nd edition, 1973.
- [67] J. Dewey, "The reflex arc concept in psychology," *Psychological Review*, vol. 3, no. 4, pp. 357–370, 1896.
- [68] W. T. Power, R. K. Clark, and R. L. McFarland, "A general feedback theory of human behavior: Part I," *Perceptual and Motor Skills*, vol. 11, no. 5, pp. 71–88, 1960.
- [69] W. T. Powers, "On Emotions and PCT: A Brief Overview," *Perceptual Control Theory: Science and Applications - A Book of Readings*, Living Control Systems, Lafayette, Colorado, 2007.
- [70] J. Gratch and S. Marsella, "A domain-independent framework for modeling emotion," *Cognitive Systems Research*, vol. 5, no. 4, pp. 269–306, 2004.
- [71] P. Gebhard, "ALMA: a layered model of affect," in *Proceedings of the Fourth International Joint Conference on*

- Autonomous Agents and Multiagent Systems*, pp. 29–36, The Netherlands, July 2005.
- [72] M. S. El-Nasr, J. Yen, and T. R. Ioerger, “Flame—fuzzy logic adaptive model of emotions,” *Autonomous Agents and Multi-Agent Systems*, vol. 3, no. 3, pp. 219–257, 2000.
- [73] C. Breazeal and B. Scassellati, “Infant-like social interactions between a robot and a human caregiver,” *Adaptive Behavior*, vol. 8, no. 1, pp. 49–74, 2000.
- [74] C. Breazeal, *Designing Sociable Robots*, MIT Press, Cambridge, MA, USA, 2002.
- [75] J. Russell, “Reading Emotions from and into Faces: Resurrecting a Dimensional-Contextual Perspective,” in *The Psychology of Facial Expression*, J. Russell and J. Fernandez-Dols, Eds., pp. 295–320, Cambridge University Press, Cambridge UK, 1997.
- [76] C. Smith and H. Scott, “10. a componential approach to the meaning of facial expressions,” *The psychology of facial expression*, vol. 229, 1997.
- [77] C. Breazeal, “Early experiments using motivations to regulate human-robot interaction,” *AAAI Fall Symposium on Emotional and Intelligent: The Tangled Knot of Cognition, Technical Report FS-98-03*, vol. 40, pp. 31–36, 1998.
- [78] C. Breazeal and L. Aryananda, “Recognition of affective communicative intent in robot-directed speech,” *Autonomous Robots*, vol. 12, no. 1, pp. 83–104, 2002.
- [79] H. Miwa, K. Itoh, H. Takanobu, and A. Takanishi, “Development of mental model for humanoid robots,” in *Proceedings of the 15th CISM-IFTOMM Symposium on Robot Design, Dynamics and Control*, Japan, September 2004.
- [80] P. Ekman and W. V. Friesen, “The repertoire of nonverbal behavior: categories, origins, usage, and coding,” *Semiotica*, vol. 1, no. 1, pp. 49–98, 1969.
- [81] H. Miwa, T. Okuchi, K. Itoh, H. Takanobu, and A. Takanishi, “A new mental model for humanoid robots for human friendly communication: introduction of learning system, mood vector and second order equations of emotion,” *IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422)*, vol. 3, pp. 3588–3593, 2003.
- [82] K. Itoh, H. Miwa, and M. Matsumoto, “Behavior Model of Humanoid Robots Based on Operant Conditioning,” in *Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots*, pp. 220–225, Tsukuba, Japan, December 2005.
- [83] R. K. Moore, “PRESENCE: a human-inspired architecture for speech-based human-machine interaction,” *IEEE Transactions on Computers*, vol. 56, no. 9, pp. 1176–1188, 2007.
- [84] G. Rizzolatti and L. J. A. R. N. Craighero, *The Mirror-Neuron System*, NCBI, Bethesda, ML, USA, 2004.
- [85] S. Ros Rodríguez and J. Hawkins, *On Intelligence*, Times Books, New York, NY, USA, 2004.
- [86] M. Wilson and G. Knoblich, “The case for motor involvement in perceiving conspecifics,” *Psychological Bulletin*, vol. 131, no. 3, pp. 460–473, 2005.
- [87] S. Saint-Aimé, B. Le-Pévédic, and D. Duhaut, “Children Recognize Emotions of Emi Companion Robot,” in *Proceedings of the 2011 IEEE International Conference on Robotics and Biomimetics*, pp. 1153–1158, Karon Beach, Thailand, December 2011.
- [88] S. Saint-Aimé, B. Le-Pévédic, and D. Duhaut, *Igrace—Emotional Computational Model for Emi Companion Robot*, InTech Education and Publishing, London, UK, 2009.
- [89] S. Saint-Aimé, C. Jost, B. Le-Pévédic, and D. Duhaut, “Dynamic Behaviour conception for Emi Companion Robot,” in *Proceedings of the ISR 2010 (41st International Symposium on Robotics) and ROBOTIK 2010 (6th German Conference on Robotics)*, pp. 1–8, Munich, Germany, June 2010.
- [90] T. H. H. Dang, S. Letellier-Zarshenas, and D. Duhaut, “Grace—generic robotic architecture to create emotions,” *Advances in Mobile Robotics: Proceedings of the Eleventh International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines*, pp. 174–181, Coimbra, Portugal, 2008.
- [91] M. S. El-Nasr, J. Yen, and T. R. Ioerger, *Autonomous Agents and Multi-Agent Systems*, vol. 3, no. 3, pp. 219–257, 2000.
- [92] F. d. Rosis, C. Pelachaud, I. Poggi, V. Carofiglio, and B. D. Carolis, “From Greta’s mind to her face: modelling the dynamics of affective states in a conversational embodied agent,” *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 81–118, 2003.
- [93] C. Adam and F. Evrard, “Donner des émotions aux agents conversationnels,” *Workshop Francophone sur les Agents Conversationnels Animés*, pp. 135–144, 2005.
- [94] Z. Kowalczyk, M. Czubenko, and T. Merta, “Interpretation and modeling of emotions in the management of autonomous robots using a control paradigm based on a scheduling variable,” *Engineering Applications of Artificial Intelligence*, vol. 91, 2020.
- [95] Z. Kowalczyk and M. Czubenko, “Emotions embodied in the SVC of an autonomous driver system,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 3744–3749, 2017.
- [96] Z. Kowalczyk and M. Czubenko, “Intelligent Decision-Making System for Autonomous Robots,” *International Journal of Applied Mathematics and Computer Science*, vol. 21, 2011.
- [97] Z. Kowalczyk and M. Czubenko, “xEmotion-obliczeniowy model emocji dedykowany dla inteligentnych systemów decyzyjnych,” *Pomiary Automatyka Robotyka*, vol. 2, no. 17, pp. 60–65, 2013.
- [98] Z. Kowalczyk and M. Czubenko, “An intelligent decision-making system for autonomous units based on the mind model,” in *Proceedings of the 23rd International Conference on Methods and Models in Automation and Robotics (MMAR)*, pp. 1–6, Miedzyzdroje, Poland, August 2018.
- [99] J. Zhang and A. J. Sharkey, “Contextual Recognition of Robot Emotions,” in *Proceedings of the 12th towards Autonomous Robotic Systems Conference (TAROS 2011)*, pp. 78–89, Sheffield, UK, September 2011.
- [100] J. Zhang and A. J. Sharkey, “Listening to Sad Music while Seeing a Happy Robot Face,” in *Proceedings of the 3rd International Conference on Social Robotics (ICSR 2011)*, pp. 173–182, Amsterdam, The Netherlands, November 2011.
- [101] J. Zhang and A. J. Sharkey, “It’s not all written on the robot’s face,” *Robotics and Autonomous Systems*, vol. 60, no. 11, pp. 1449–1456, 2012.
- [102] J. Zhang, *Contextual Recognition of Robot Emotions*, PhD Thesis, University of Sheffield, Sheffield, UK, 2013.
- [103] S. Ojha, J. Vitale, and M. A. Williams, “Computational emotion models: a thematic review,” *International Journal of Social Robotics*, vol. 13, no. 6, pp. 1253–1279, 2021.
- [104] D. Hu, H. Wang, F. Deng et al., “A facial emotion cognition and training system for autism rehabilitation based on deep learning and mobile technology,” *Basic and Clinical Pharmacology and Toxicology*, vol. 125, p. 46, 2019.
- [105] G. Wan, F. Deng, Z. Jiang et al., “FECTS: A Facial Emotion Cognition and Training System for Chinese Children with

- Autism Spectrum Disorder,” *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 9213526, 21 pages, 2022.
- [106] S. Lin, J. Su, S. Song, and J. Zhang, “An event-triggered low-cost tactile perception system for social robot’s whole body interaction,” *IEEE Access*, vol. 9, pp. 80986–80995, 2021.
- [107] X. Li, H. Zhong, B. Zhang, and J. Zhang, “A general Chinese chatbot based on deep learning and its’ application for children with ASD,” *International Journal of Machine Learning and Computing*, vol. 10, no. 4, pp. 519–526, 2020.
- [108] B. Zhang, L. Zhou, and S. Song, “Image captioning in Chinese and its application for children with autism spectrum disorder,” in *Proceedings of the 2020 12th International Conference on Machine Learning and Computing*, pp. 426–432, Shenzhen China, February 2020.
- [109] Z. Li, Y. Mu, Z. Sun, S. Song, J. Su, and J. Zhang, “Intention understanding in human-robot interaction based on visual-NLP semantics,” special issue of intelligence and safety for humanoid robots: design, control, and applications,” *Frontiers in Neurorobotics*, vol. 14, p. 121, 2021.
- [110] F. Deng, Y. Zhou, S. Song et al., “Say what you are looking at: an attention-based interactive system for autistic children,” *Applied Sciences*, vol. 11, no. 16, p. 7426, 2021.
- [111] X. Cai, Z. Yan, F. Duan, D. Hu, and J. Zhang, “Lightweight Convolution Neural Network Based on Feature Concatenate for Facial Expression Recognition,” *Intelligent Computing in Engineering*, vol. 1125, pp. 1141–1148, 2019.
- [112] K. Itoh, H. Miwa, and Y. Nukariya, “Behavior generation of humanoid robots depending on mood,” in *Proceedings of the 9th International Conference on Intelligent Autonomous Systems (IAS-9)*, pp. 965–972, IOS Press, Tokyo, Japan, March 2006.
- [113] R. R. McCrae and P. T. Costa, “Validation of the five-factor model of personality across instruments and observers,” *Journal of personality and social psychology*, vol. 52, no. 1, 1987.
- [114] P. T. Costa and R. R. McCrae, “Normal personality assessment in clinical practice: the NEO Personality Inventory,” *Psychological Assessment*, vol. 4, no. 1, pp. 5–13, 1992.
- [115] L. Robert, “Personality in the human robot interaction literature: a review and brief critique personality in the human robot interaction literature: a review and brief critique,” in *Proceedings of the 24th Americas Conference on Information Systems*, pp. 16–18, New Orleans, LA, USA, August 2018.
- [116] J. Veroff, “Contextual determinants of personality,” *Personality and Social Psychology Bulletin*, vol. 9, no. 3, pp. 331–343, 1983.
- [117] M. Joosse, M. Lohse, J. G. Perez, and V. Evers, “What You Do Is Who You Are: The Role of Task Context in Perceived Social Robot Personality,” in *Proceedings of the 2013 IEEE International Conference on Robotics and Automation*, pp. 2134–2139, Karlsruhe, Germany, May 2013.
- [118] A. Andriella, H. Siqueira, D. Fu et al., “Do i have a personality? endowing care robots with context-dependent personality traits,” *International Journal of Social Robotics*, vol. 13, no. 8, pp. 2081–2102, 2020.
- [119] C. Esterwood and L. P. Robert, “A systematic review of human and robot personality in health care human-robot interaction,” *Frontiers in Robotics and AI*, vol. 8, Article ID 748246, 2021.
- [120] A. Mehrabian, “Analysis of the big-five personality factors in terms of the PAD temperament model,” *Australian Journal of Psychology*, vol. 48, no. 2, pp. 86–92, 1996.
- [121] E. M. Tong, G. D. Bishop, H. C. Enkelmann et al., “The role of the Big Five in appraisals,” *Personality and Individual Differences*, vol. 41, no. 3, pp. 513–523, 2006.
- [122] N. Masuyama, C. K. Loo, and M. Seera, “Personality affected robotic emotional model with associative memory for human-robot interaction,” *Neurocomputing*, vol. 272, pp. 213–225, 2018.
- [123] O. Rudovic, J. Lee, M. Dai, B. Schuller, and R. W. Picard, “Personalized machine learning for robot perception of affect and engagement in autism therapy,” *Science Robotics*, vol. 3, no. 19, Article ID eaa06760, 2018.
- [124] E. Schopler, M. E. Van Bourgondien, G. J. Wellman, and S. R. Love, *Childhood Autism Rating Scale™*, Cognitive Centre, Saravanampatti, Coimbatore, India, 2nd edition, 2011.
- [125] C. Lord, M. Rutter, and P. DiLavore, *Autism diagnostic observation schedule–(ADOS-2)*, Western Psychological Corporation, Los Angeles, CA, USA, 2012.
- [126] M. Rutter, A. Le-Couteur, and C. Lord, *Autism Diagnostic Interview-Revised*, Western Psychological Services, Los Angeles, CA, USA, 2003.
- [127] M. F. Folstein, S. E. Folstein, and P. R. McHugh, “Minimal state”: a practical method for grading the cognitive state of patients for the clinician,” *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, 1975.

Research Article

Path Planning Algorithm for Unmanned Surface Vessel Based on Multiobjective Reinforcement Learning

Caipei Yang,¹ Yingqi Zhao,¹ Xuan Cai,² Wei Wei,² Xingxing Feng,² and Kaibo Zhou¹ 

¹MOE Key Laboratory of Image Information Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China

²Wuhan Second Ship Design and Research Institute, Wuhan 430205, China

Correspondence should be addressed to Kaibo Zhou; zhoukb@hust.edu.cn

Received 5 July 2022; Revised 12 October 2022; Accepted 20 January 2023; Published 15 February 2023

Academic Editor: Abdul Rehman Javed

Copyright © 2023 Caipei Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is challenging to perform path planning tasks in complex marine environments as the unmanned surface vessel approaches the goal while avoiding obstacles. However, the conflict between the two subtarget tasks of obstacle avoidance and goal approaching makes the path planning difficult. Thus, a path planning method for unmanned surface vessel based on multiobjective reinforcement learning is proposed under the complex environment with high randomness and multiple dynamic obstacles. Firstly, the path planning scene is set as the main scene, and the two subtarget scenes including obstacle avoidance and goal approaching are divided from it. The action selection strategy in each subtarget scene is trained through the double deep Q-network with prioritized experience replay. A multiobjective reinforcement learning framework based on ensemble learning is further designed for policy integration in the main scene. Finally, by selecting the strategy from subtarget scenes in the designed framework, an optimized action selection strategy is trained and used for the action decision of the agent in the main scene. Compared with traditional value-based reinforcement learning methods, the proposed method achieves a 93% success rate in path planning in simulation scenes. Furthermore, the average length of the paths planned by the proposed method is 3.28% and 1.97% shorter than that of PER-DDQN and dueling DQN, respectively.

1. Introduction

In ocean exploration, the competition among countries to protect marine territorial sovereignty and develop marine resources has become increasingly fierce. The unmanned surface vessel (USV), as a kind of vessel with high autonomy, has broad application prospects in the field of ocean exploration. As one of the current research hotspots, the path planning of USV faces many challenges, including unknown environment, perceptual uncertainty, and dynamic obstacles [1–3]. The USV path planning is aimed to obtain a collision-free path under specific circumstances. It can be divided into two subtarget tasks, such as goal approaching and obstacle avoidance. The goal approaching method helps the USV reach the destination, focusing on reducing path length and travel time. The obstacle avoidance method makes the USV conduct real-time collision avoidance through a series of decisions [4].

Traditional path planning methods perform well in simple known static environments and reach a destination while avoiding obstacles [5–9]. But there are still major deficiencies in the exploration and decision-making capabilities of algorithms in complex environments, failing to guarantee the success rate and environmental adaptability. Currently, the deep reinforcement learning (DRL) methods have advantages in unknown environment exploration and real-time action decision making in path planning problems [10, 11]. Therefore, the use of DRL methods to solve the path planning problem has become one of the new research directions [12]. For example, Tai et al. used radar observations and target positions as inputs and applied DRL methods to path planning tasks for the first time [13]. The agent uses the discrete control commands generated by the algorithm to avoid obstacles in the indoor mobile environment. Chen et al. proposed an intelligent collision

avoidance algorithm with DRL improving the path quality compared with optimal reciprocal collision avoidance (ORCA) [14]. Chen et al. constructed the interaction model between the agent and the obstacle, providing the basis for the reinforcement learning strategy of the agent's path planning in complex dynamic environment [15]. Thus, it is effective to use DRL algorithms for goal approaching and dynamic obstacle avoidance.

However, the path length inevitably increases in the obstacle avoidance process, which conflicts with the requirement of destination reaching for the goal approaching subtarget task. Therefore, it is difficult for a single optimization strategy to simultaneously achieve these subtarget tasks. Recently, intelligence computing algorithms have been widely used in related fields [16–19]. A more comprehensive model can be obtained in ensemble learning by combining multiple weak learners [20]. Inspired by the idea of integrated learning, a multiobjective reinforcement learning architecture is designed to trade off these subtarget tasks. There is a need to investigate the USV path planning based on multiobjective reinforcement learning.

Main contributions in this paper can be summarized as follows:

- (1) Based on the main scene of path planning considering random goals and multiple dynamic obstacles, the dynamic obstacle avoidance subtarget scene and the goal approaching subtarget scene are constructed. The double deep Q-network with prioritized experience replay (PER-DDQN) is applied to the action decision of USV in two scenes, respectively.
- (2) A multiobjective reinforcement learning architecture based on ensemble learning is designed, optimizing the multiobjective policy integration method in the USV path planning task.
- (3) A USV path planning algorithm based on multiobjective reinforcement learning is proposed, improving the success rate of USV path planning tasks and shortening the planned path length in the complex environment.

The rest of this paper is organized as follows. The theoretical background of the PER-DDQN and the multiobjective reinforcement learning method is introduced in Section 2. The proposed algorithm is introduced in Section 3. Simulation experiments and results are presented in Section 4. Discussion is given in Section 5, and Section 6 concludes this paper.

2. Related Work

2.1. Q-Learning. The Q-learning algorithm is a value-based reinforcement learning algorithm [21]. A Q-value table is built and updated in the Q-learning algorithm. Each action is selected with the greatest benefit based on the Q-value. The maximum Q-value of the next state is used to estimate the Q-value of the current state. The update formula is as follows:

$$Q(s, a) = Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s, a) - Q(s, a)], \quad (1)$$

where $Q(s, a)$ denote agent's expectation of reward for performing action a in state s . α represents the learning rate and γ represents the discount factor. The reward obtained by the agent after performing action a is r , and the state is changed to s' . $Q(s, a)$ denote agent's expectation of reward for performing action a' in state s' .

2.2. Deep Q-Network. To address the curse of dimensionality in high-dimensional state spaces, Mnih et al. used a neural network with θ to approximate the Q-value: $Q(s, a; \theta) \approx Q(s, a)$ [22]. DQNs are optimized by reducing and minimizing $L_i(\theta_i) = E_{s,a,r,s'} [(y_i^{\text{DQN}} - Q(s, a; \theta_i))^2]$ at each iteration i , with target $y_i^{\text{DQN}} = r + \gamma \max_{a'} Q(s, a; \theta_i^-)$. Here, θ_i^- are the parameters of a target network that is frozen for a number of iterations while updating the online network $Q(s, a; \theta_i)$ by gradient descent. The action a is chosen from $Q(s, a; \theta_i)$ by an action selector, which typically implements an ϵ -greedy policy that selects the action that maximizes the Q-value with a probability of $1-\epsilon$ and chooses randomly with a probability of ϵ .

2.3. Experience Replay. Online reinforcement learning (RL) agents incrementally update their parameters (of the policy, value function or model) while they observe a stream of experience [23]. Because the agent discards experience after one update in simple reinforcement learning, rare valid experience is underutilized. At the same time, there is a substantial correlation between neighbouring experiences, which is not favourable to model training. By storing experiences in replay memory, experience replay can effectively solve the above problems. It becomes possible to break the temporal correlations by mixing more and less recent experience for the updates [24].

2.4. Related Literature. Value function-based DRL algorithm uses deep neural network to approximate value function or action value function and uses temporal difference or Q-learning, respectively, to update the value function or action value function. Many scholars use DRL methods based on value functions, including DQN algorithm and some improved variant algorithms, to motivate robots or other agents to obtain optimal paths [25–27]. Additionally, with the introduction of the strategy gradient method, DRL based on strategy gradient is used in robot path planning, such as A3C [28], DDPG [29], TRPO [30], and PPO [31]. When it comes to agent data control and management, blockchain hyperledger fabric is one of the practical technologies [32, 33]. We have briefly summarized some of the recent literature, as shown in Table 1.

3. Methodology

When the USV performs a mission in a complicated marine environment with various dynamic impediments, it needs to arrive at its destination without colliding with the obstacles.

TABLE 1: Related literature.

	Study title	Approach	Merit	Limitations	Ref
DRL based on value function	An improved algorithm of robot path planning in complex environment based on double DQN	Double DQN	The problem of lacking experiments is solved by redefining the initialization of the robot and the reward function for the free position	Slow convergence speed of the algorithm	[25]
	The USV path planning of dueling DQN algorithm based on tree sampling mechanism	Dueling DQN	The algorithm can identify and avoid static obstacles in the environment and realize autonomous navigation in complex environments	Internal connection between the state-action pairs is not strong enough	[26]
	Tactical UAV path optimization under radar threat using deep reinforcement learning	DQN-PER	Alleviates the sparse reward problem	Overvaluation of the action-state value	[27]
DRL based on strategy gradient	Advanced double layered multi-agent systems based on A3C in real-time path planning	A3C	The correlation between state distribution samples is eliminated, and the sample storage mode of experience playback mechanism is replaced	Convergence to local optimal strategy	[28]
	The path-planning algorithm of unmanned ship based on DDPG	DDPG	The algorithm can be applied to continuous state space and action space	Sensitive to hyperparameters	[29]
	Hindsight trust region policy optimization	TRPO	The algorithm can choose a more appropriate step length during training	Large environments and policies are prone to large errors	[30]
	PPO-based reinforcement learning for UAV navigation in urban environments	PPO	The algorithm has better data efficiency and robustness	The difference between the old and new policies cannot be too large with each update	[31]

It is necessary to create a model that can select appropriate actions in different states in order to achieve dynamic obstacle avoidance and goal approaching.

3.1. PER-DDQN. The PER-DDQN improves the learning effect and the learning speed by introducing the DDQN and priority experience replay. Two Q-networks are used in DDQN to eliminate the bias caused by the greedy policy [34]. The current Q-network is used to calculate the action corresponding to the maximum Q-value, and the target Q-network is used to calculate the target Q-value corresponding to the maximum action. Prioritized experience replay is a stochastic sampling method that interpolates between pure greedy prioritization and uniform random sampling [35]. The probability of being sampled is monotonic in a transition's priority, while guaranteeing a nonzero probability even for the lowest priority transition. The probability of sampling transition i is defined as

$$P(i) = \frac{p_i^\alpha}{\sum_k p_k^\alpha}, \quad (2)$$

where i is the priority of transition. The exponent α determines how much prioritization is used, with $\alpha=0$ corresponding to the uniform case. In the actual process, all samples can be divided into n intervals, and uniform sampling is performed in each interval. The PER-DDQN is

used for the action decision of the agent in the constructed scene. The flowchart of the algorithm is shown in Figure 1.

3.2. Framework for Multiobjective Reinforcement Learning.

The path planning task of the USV includes two subtarget tasks, such as dynamic obstacle avoidance and goal approaching. The traditional reinforcement learning architecture for a single task is no longer appropriate. A multiobjective reinforcement learning architecture is built for policy learning and ensemble in the main scene of path planning, inspired by ensemble learning.

The fundamental principle of ensemble learning is to integrate the learning results of numerous weak models to produce better overall results, which can be classified into bagging, boosting, and stacking. The sample training set is sampled with replacement in the bagging method, yielding T independent sample sampling sets. T weak learners are trained from T sample sets. Weighted average, voting, and other strategy integration approaches are employed to provide final decision results [36]. Figure 2 depicts the flowchart.

Corresponding to weak learners in ensemble learning, the designed multiobjective reinforcement learning architecture leverages subagents for training in subtarget scenes. Different from traditional integration methods, the proposed method uses a main agent based on the reinforcement learning method for policy integration. According to the

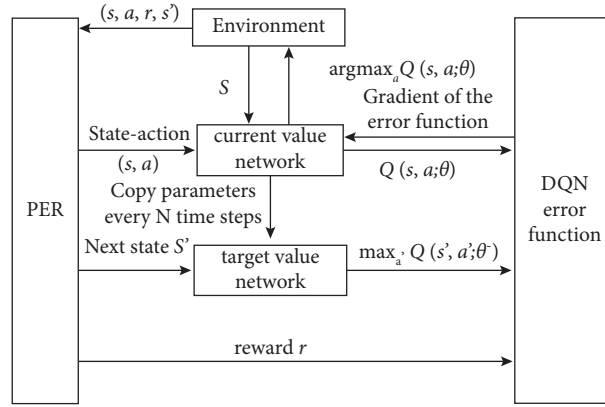


FIGURE 1: Strategy iteration and optimization based on PER-DDQN reinforcement learning algorithm.

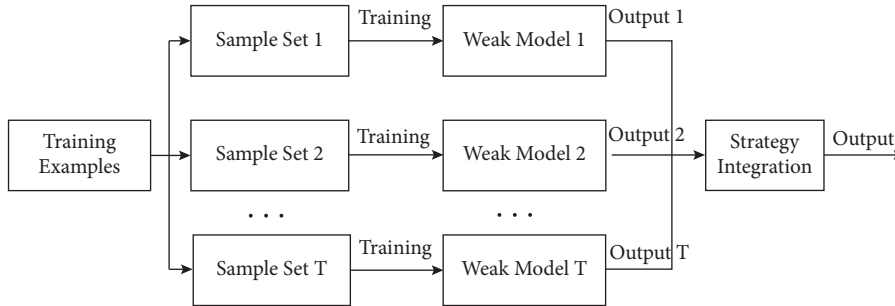


FIGURE 2: Flowchart of the bagging algorithm.

environmental state of the main scene, the main agent selects the strategy of the subagent in the corresponding state of the subscene and makes a decision. The designed multiobjective reinforcement learning architecture is shown in Figure 3.

3.3. The Proposed Approach. The PER-DDQN algorithm is combined with the designed architecture for the constructed path planning scene, and a USV path planning algorithm based on multiobjective reinforcement learning is proposed. Figure 4 depicts the overall process of the proposed method.

Step 1. The subagents in each subtarget scene are trained using the PER-DDQN algorithm, and the strategies of each subagent are saved.

Step 2. In the constructed path planning main scene, the main agent is trained by the PER-DDQN method. The main agent selects subagent according to the current environment state and gives the actions according to the strategy of the selected subagent in this state.

Step 3. The main agent executes actions of the selected subagent, generating and storing experience for the main agent to learn from.

4. Simulation Experiments

The main scene, dynamic obstacle avoidance subtarget scene, and goal approaching subtarget scene are built in Unity3D to verify the effectiveness of the proposed method. The settings for scenario conditions and reinforcement learning parameters are provided separately. Algorithms were written by Python 3.8 and processed by a server with a RAM (64G) and a CPU (Intel Core i9-11900K).

4.1. Scene Building. The main scene of path planning considering random goal and multiple dynamic obstacles is generated on a two-dimensional plane, as illustrated in Figure 5, to represent the complicated marine environment.

The dynamic obstacle avoidance subtarget scene is built on the basis of the main scene, as shown in Figure 6, to focus on the dynamic obstacle avoidance subtarget. The agent does not need to consider the problem of goal approaching and instead attempts to travel through the obstacle region without colliding. It is deemed effective obstacle avoidance when the agent's ordinate is larger than the ordinate of all

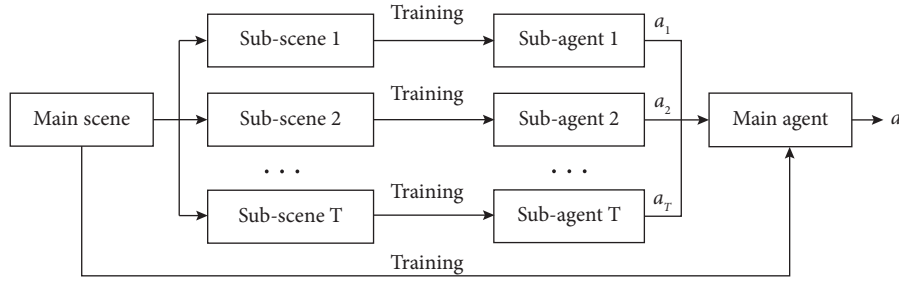


FIGURE 3: Proposed framework for multiobjective reinforcement learning.

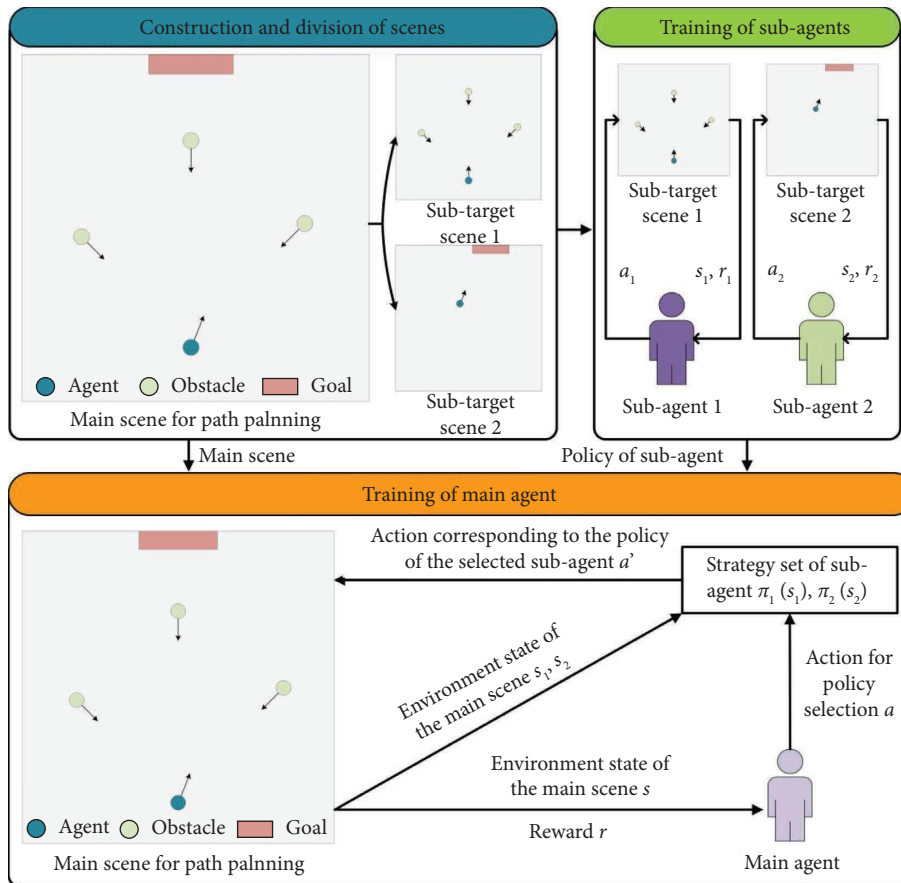


FIGURE 4: The overall process of the path planning algorithm for USV based on multiobjective reinforcement learning.

obstacles. When a collision occurs, it is regarded as obstacle avoidance failure.

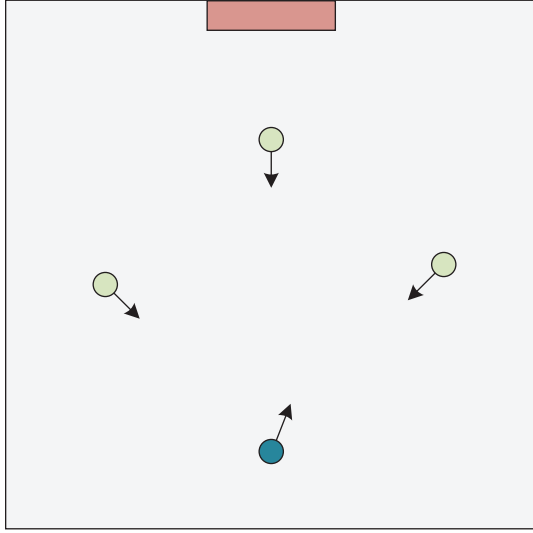
The goal approaching subtarget scene is built on the basis of the main scene, as shown in Figure 7, to focus on the goal approaching subtarget. Dynamic obstacles are removed, and the only learning objective is to approach the goal.

4.2. Simulation Setup

4.2.1. *Initial Conditions.* The initial conditions of agents, dynamic obstacles, and goals in the main scene and each subtarget scene (dynamic obstacle avoidance and goal approaching) are set to random values to ensure that the

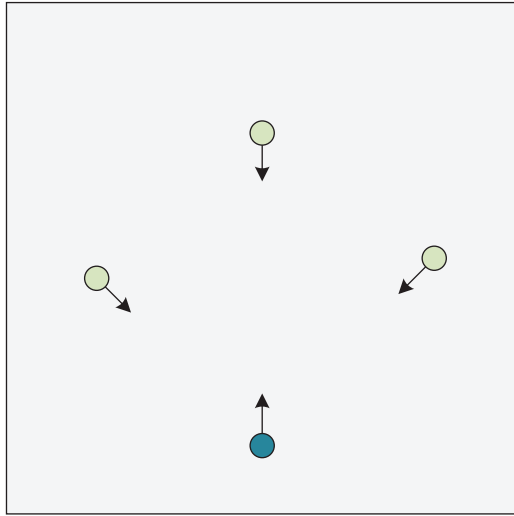
training model generalizes and meets the actual application requirements.

(1) *Dynamic Obstacle Avoidance Subtarget Scene.* For dynamic obstacles, set its radius to 0.5 m, the maximum speed to 1 m/s, and the quantity to 3. The states of three obstacles are set using the A* path planning algorithm and the ORCA dynamic obstacle avoidance algorithm to avoid mutual collision. The coordinates of dynamic obstacles' starting points are randomly selected in square areas centered on (0 m, 6 m), (5 m, 5 m), and (-5 m, 5 m), respectively. Also, the coordinates of dynamic obstacles' end points are randomly selected in square areas centered on (0 m, -6 m), (-5 m, -5 m), and (5 m, 5 m). The area of each area is 4 m².



■ Goal
● Agent (USV)
● Dynamic Obstacle

FIGURE 5: Main scene for USV path planning.



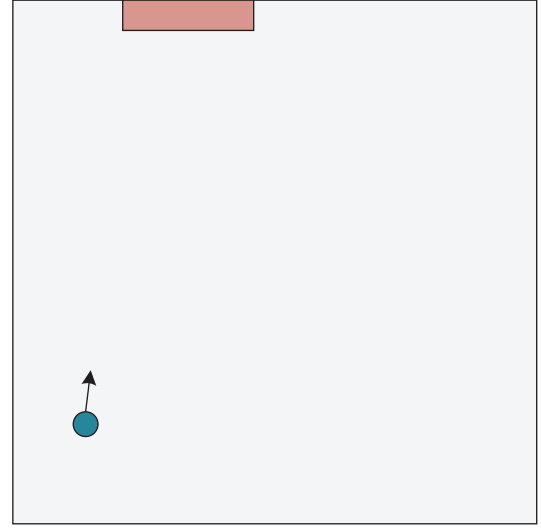
● Agent (USV)
● Dynamic Obstacles

FIGURE 6: Dynamic obstacle avoidance subtarget scene.

For the agent, set its radius to 0.5 m and the maximum speed to 1 m/s. The agent's initial abscissa and ordinate are chosen randomly from the range $[-2 \text{ m}, 2 \text{ m}]$, $[-4 \text{ m}, -8 \text{ m}]$.

(2) *Goal Approaching Subtarget Scene.* The goal is a rectangle with a length of 5 m and a width of 1 m. The initial abscissa of the target is randomly selected within the range of $[-5 \text{ m}, 5 \text{ m}]$. The initial abscissa and ordinate of the agent are chosen randomly from the range $[-6 \text{ m}, 6 \text{ m}]$, $[-8 \text{ m}, 4 \text{ m}]$.

(3) *Main Scene.* The initial condition of the goal is consistent with the goal in the goal approaching subtarget scene, and the motion parameters of the dynamic obstacle are



■ Goal
● Agent (USV)

FIGURE 7: Goal approaching subtarget scene.

consistent with dynamic obstacles in the dynamic obstacle avoidance subtarget scene. The agent's initial abscissa is chosen randomly from the range $[-2 \text{ m}, 2 \text{ m}]$, $[-4 \text{ m}, -8 \text{ m}]$, and the initial ordinate is set to -8 m .

4.2.2. *Reinforcement Learning Parameter.* The reinforcement learning settings for the agent in each scene, such as the action space, state, and rewards, are set as follows.

(1) *Action Space Setting.* The agent's action space is set to 5 directions divided evenly into in the main scene and each subscene, as shown in Figure 8, to reduce the training cost.

(2) *State Settings.* The states of the agent ($s_{1,t}$ and $s_{2,t}$) are set as equations (3) and (4) in two subtarget scenes (dynamic obstacle avoidance and goal approaching):

$$s_{1,t} = (s_{self1}, s_{obs1}, s_{obs2}, s_{obs3}), \quad (3)$$

$$s_{2,t} = (s_{self2}, s_{tgt}), \quad (4)$$

where S_{self1} and S_{self2} are the states of the agent in two subtarget scenes, represented by agent's positions, respectively, at $t-2$, $t-1$, t . S_{obs1} , S_{obs2} , and S_{obs3} are the states of the obstacles in dynamic obstacle avoidance subtarget scene, represented by obstacles' positions, respectively, at $t-2$, $t-1$, t . S_{tgt} is the state of the goal in goal approaching subtarget scene, represented by goal's position, respectively, at t .

The state of agent s_t is set as equation (5) in the main scene:

$$s_t = (s'_{self}, s'_{obs1}, s'_{obs2}, s'_{obs3}, s'_{tgt}), \quad (5)$$

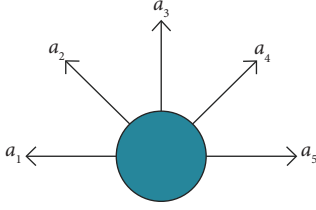


FIGURE 8: Action space of agent.

where s'_{obs1} , s'_{obs2} , s'_{obs3} , and s'_{self} are the states of the obstacles and the agent in the main scene, represented by their positions, respectively, at $t-2$, $t-1$, t . s'_{gt} is the state of the goal in the main scene, represented by goal's position, respectively, at t . In dynamic obstacle avoidance subtarget scene, the dimensions of action space and observation space are 5 and 8, respectively. In goal approaching subtarget scene, the dimensions of action space and observation space are 5 and 3, respectively. In dynamic obstacle avoidance subtarget scene, the dimensions of action space and observation space are 5 and 9, respectively.

(3) *Reward Setting*. The rewards of the agent ($r_{1,t}$, $r_{2,t}$, and r_t) are set as equations (6)–(8) in two subtarget scenes and the main scene:

$$r_{1,t} = \begin{cases} -0.5 & \text{if: } 1 < \text{obs_dist} < 1.05, \\ -2 & \text{obs_dist} = 1, \\ 1.5 & y_t > y_{i,t}, \forall i = 1, 2, 3, \\ 0.3 * (y_t - y_{t-1}) - 0.015 & \text{else,} \end{cases} \quad (6)$$

$$r_{2,t} = \begin{cases} 1.5 & \text{if: target_collided,} \\ 0.2 * (\text{pre_dist}) - 0.005 & \text{else,} \end{cases} \quad (7)$$

$$r_t = \begin{cases} -0.5 & \text{if: } 1 < \text{obs_dist} < 1.05, \\ -1.5 & \text{obs_dist} = 1, \\ 1.5 & y_t > y_{i,t}, \forall i = 1, 2, 3, \\ 0.5 * (\text{pre_dist} - \text{dist}) - 0.005, & \end{cases} \quad (8)$$

where obs_dist represents the distance between the agent and the dynamic obstacle at time t . y_t and y_{t-1} are the ordinates of the agent at time t and time $t-1$, respectively. target_collided indicates whether the agent is in contact with the goal, and dist and pre_dist are the distances between the agent and the goal at time t and time $t-1$, respectively. The training parameters of the reinforcement learning algorithm in each scene are shown in Table 2.

4.3. Result Analysis. After 800 times of training in the dynamic obstacle avoidance subtarget scene, the two samples are shown in Figure 9. The agent's obstacle avoidance strategy is slightly different in different scenes. The dynamic obstacles are evenly dispersed in front of the agent, as indicated in Figure 9(a), and the collision risk is substantial.

The agent chooses to move to the right, avoiding the range where obstacles might congregate. The dynamic obstacles are concentrated at the agent's front right, as shown in Figure 9(b), and the agent chooses to go straight at the start. When there is a risk of collision, the agent turns left to avoid obstacles urgently.

After 800 times of training in the goal approaching subtarget scene, the two samples are shown in Figure 10. When the goal is in front of the agent, as shown in Figure 10(a), the agent continues to adjust at the beginning and end of the path while moving forward in the middle. When the goal is far from the front of the agent, as shown in Figure 10(b), the agent remains adjusted throughout. The results show that the agent can rapidly approach the goal under various initial conditions without the interference of dynamic obstacles.

After 800 times of training in the main scene, the two samples are shown in Figure 11. As shown in Figure 11(a), the obstacles are distributed in front of the agent. At the same time, the target is far from the front of the agent. The agent chooses to move sideways quickly after going straight through the obstacle area in the initial stage to approach the goal. As shown in Figure 11(b), the obstacles are evenly distributed in front of the agent. At the same time, the target is near the front of the agent. The agent chooses to go straight and dynamically avoid collision in the obstacle area.

The experimental results show that by dynamically selecting the strategy of subagents, the main agent can avoid obstacles and approach the goal in various scenes to accomplish the path planning task well. Therefore, the effectiveness of the proposed method has been verified.

5. Discussion

To verify the effectiveness of the proposed framework on strategy integration, a comparison is made between reinforcement learning methods that use integration methods such as linear voting method and rank voting method and our method. At the same time, the proposed method is compared with A^* + ORCA and the path planning algorithm based on single-objective reinforcement learning to demonstrate the advantages of the proposed method in path planning tasks.

5.1. Comparison with Other Ensemble Learning Algorithms.

In the linear voting method, the Q -value of each action in the main scene is the normalized sum of the Q -values in the corresponding states of each subscene. In the rank voting method, the rank of each action in the main scene is the sum of the ranks of the corresponding states of each subscene. In these methods, the subagents and their strategies are consistent with the proposed method. The performance indicators of these methods in the results of 100 random experiments are shown in Table 3.

The rank voting method has the worst integration effect and the lowest success rate. Compared with the rank voting method, the linear voting method considerably enhances the success rate by keeping the path length from increasing. The path length of the proposed method is slightly longer than

TABLE 2: Training parameters for reinforcement learning.

Scene	Learning rate	Batch_size	Discount factor	FC structure
Dynamic obstacle avoidance	0.0005	256	0.99	[64, 64, 32]
Goal approaching	0.001	256	0.99	[32, 32, 16]
Main scene	0.0005	256	0.99	[32, 32, 16]

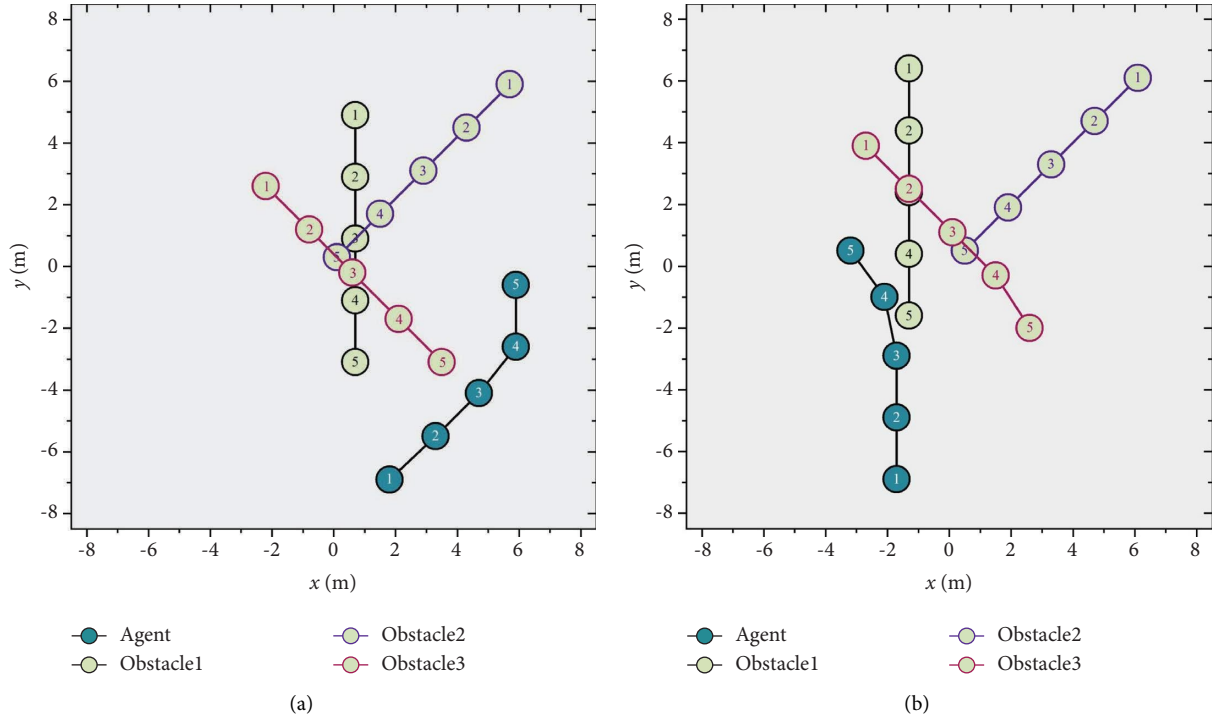


FIGURE 9: Experimental results of the agent's dynamic obstacle avoidance subscene.

that of the other two methods, but the success rate of goal approaching and dynamic obstacles avoidance are higher. The proposed method has the best overall performance.

Three random samples of the path planning results of the three methods in the same environment are shown in Figure 12. "CA" represents the main agent to choose the strategy in the dynamic obstacle avoidance subtarget scene. "TA" represents the main agent to choose the strategy in the goal approaching subtarget scene. Each agent makes better decision in low-complexity environments to avoid obstacles and approach the goal. In a more complicated context, however, the agents using the traditional ensemble method face the issue of disordered decision making. The decision strategy of the obstacle avoidance agent is selected in the initial stage of the path to maximize the success rate of obstacle avoidance. When approaching the goal, the decision strategy of the goal approaching agent is selected to maximize the success rate of goal approaching. The proposed method has a greater success rate of path planning in varied situations than the other two ensemble methods, demonstrating the superiority of the proposed ensemble learning architecture over the traditional ensemble methods.

5.2. Comparison with Other Path Planning Algorithms. The training times of PER-DDQN are 2400 times, and other hyperparameters are consistent with the method's parameter

setting in the main scene. The performance indicators of these methods in the results of 100 random experiments are shown in Table 4.

The assumptions of the ORCA in the decision-making process are inconsistent with the requirements of dynamic obstacle avoidance in practical applications. Therefore, the success rate of agent using the A^* + ORCA is low. DDQN algorithm solved the problem of overestimation of action value function in Q-learning. On this basis, PER-DDQN uses priority sampling to accelerate the convergence speed of the algorithm, and dueling DQN uses the competitive architecture to estimate the value function more precisely. They perform well in the constructed scenes. Our approach combines the strengths of reinforcement learning with ensemble learning. The experimental results show that the method proposed in this paper has the best overall performance when considering path length and success rate.

Four random samples of the path planning results of the four methods in the same environment are shown in Figure 13. The policies provided by the A^* + ORCA method are not sufficient for the agent to always avoid obstacles. The policies provided by dueling DQN are conservative, and there may be detours. The policies provided by the PER-DDQN are not mature enough in dealing with conflicts between subtarget tasks. Many problems still exist such as long planning path, failure to avoid obstacles, and

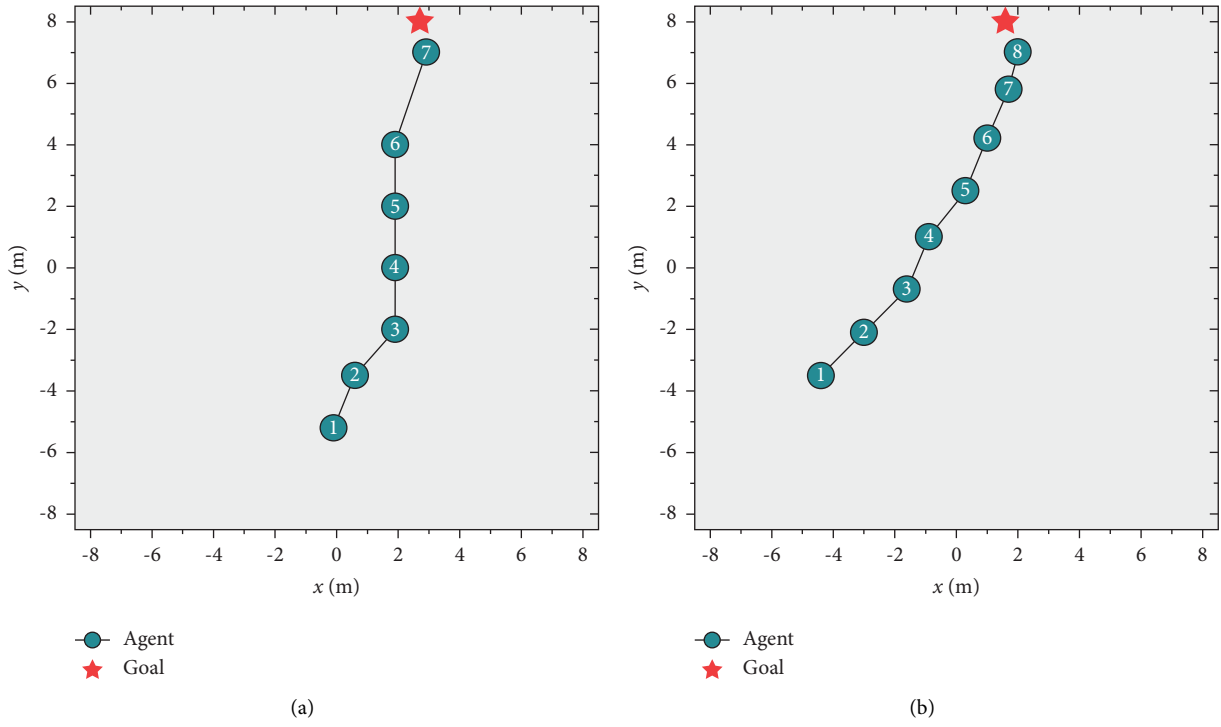


FIGURE 10: Experimental results of the agent's target approach subscene.

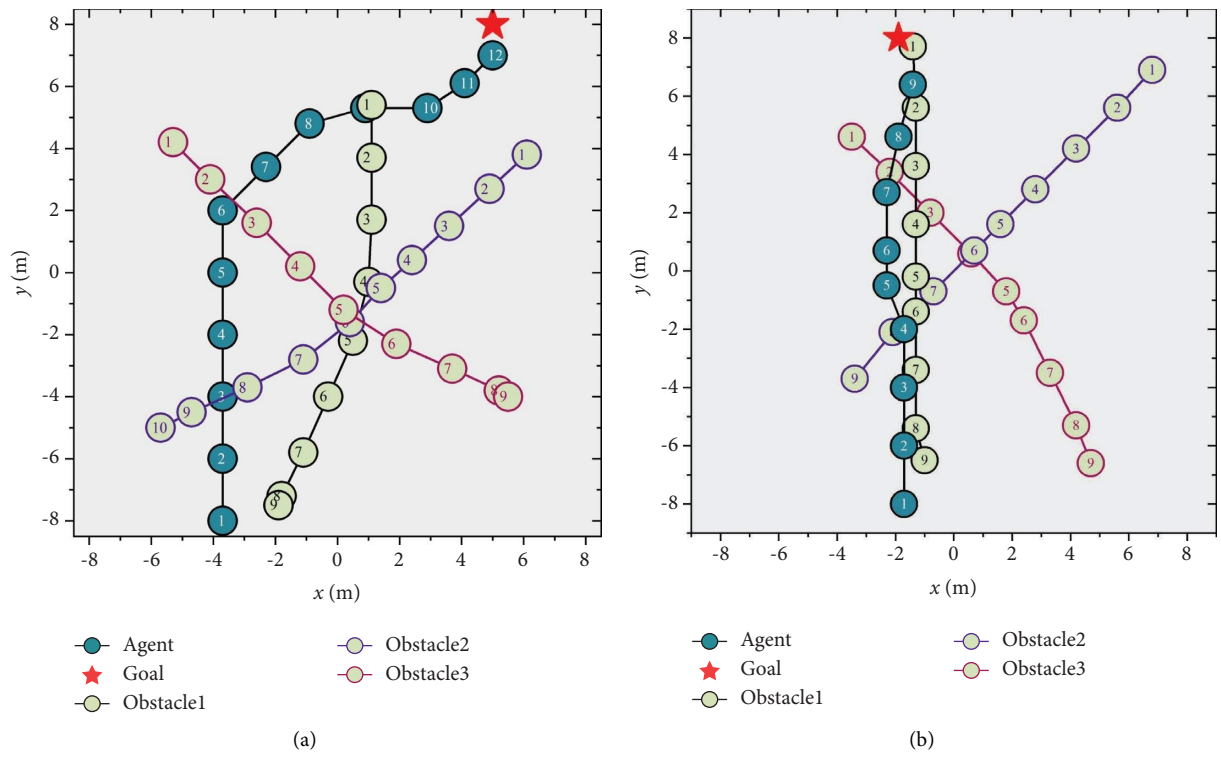


FIGURE 11: Experiment results of the main scene of agent path planning.

TABLE 3: Performance of various ensemble learning methods.

Method of policy integration	Success rate of path planning (%)	Success rate of goal approaching (%)	Success rate of dynamic obstacle avoidance (%)	Length of the path (m)
Rank voting	52	55	91	17.68
Linear voting	57	63	94	16.98
Proposed method	93	96	97	17.96

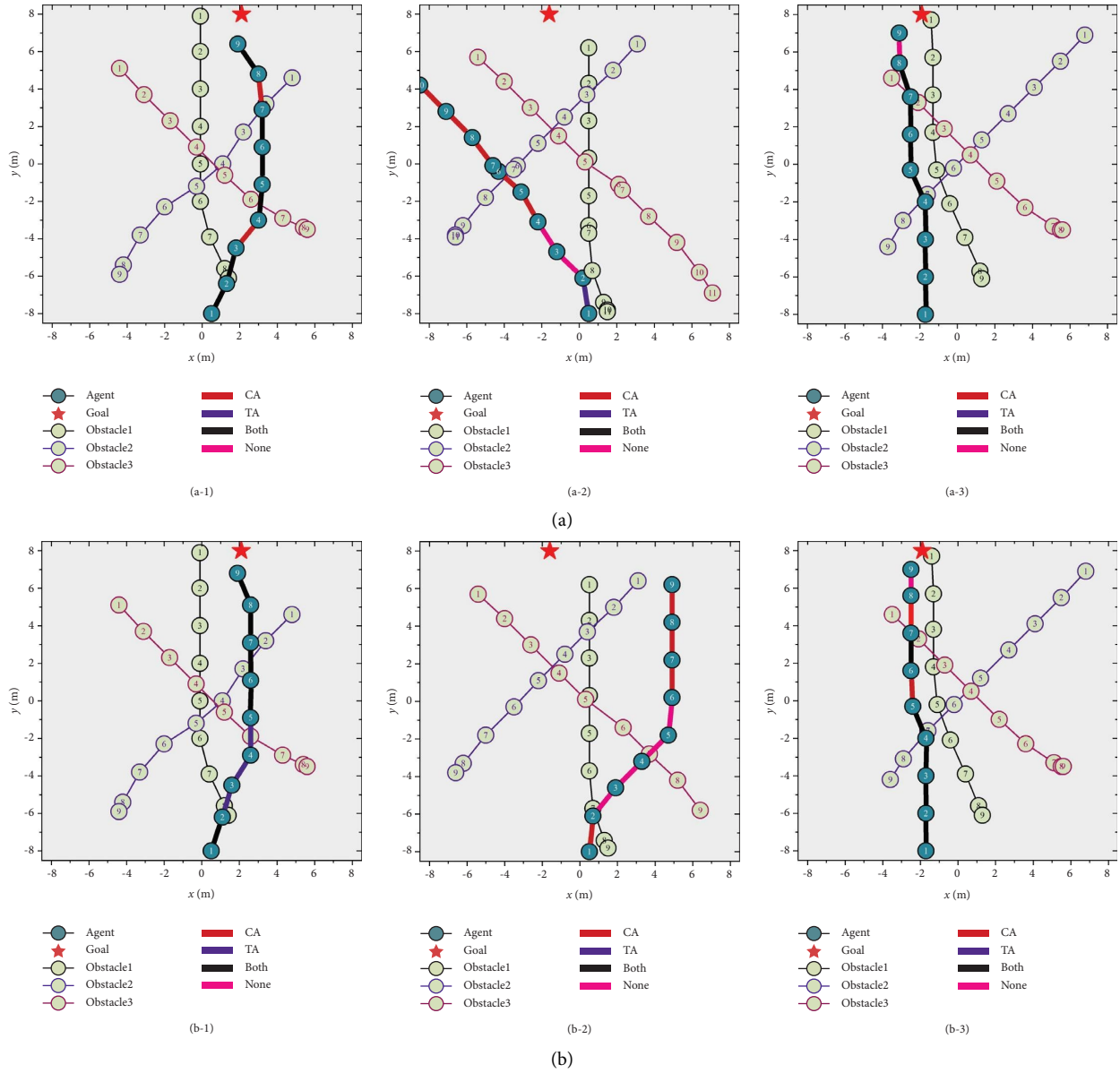


FIGURE 12: Continued.

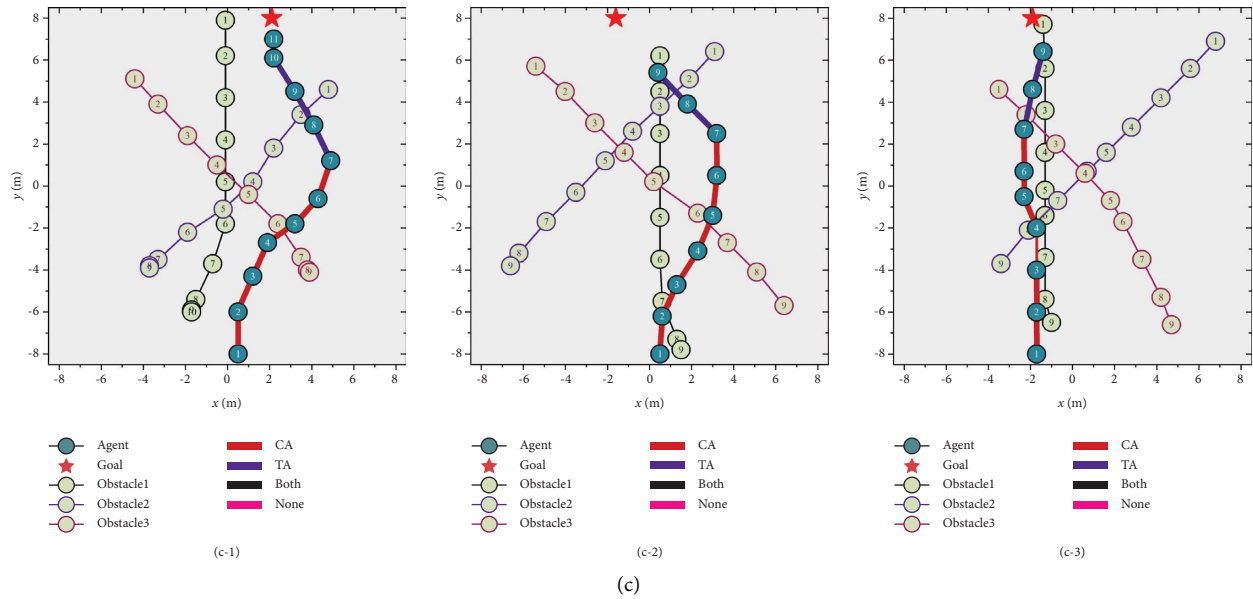


FIGURE 12: Path planning results of various ensemble learning methods in the same environment: (a) rank voting, (b) linear voting, and (c) proposed method.

TABLE 4: Performance of various path planning methods.

Method of path planning	Success rate of path planning (%)	Success rate of goal approaching (%)	Success rate of dynamic obstacle avoidance (%)	Length of the path (m)
A* + ORCA	42	100	42	16.43
Dueling DQN	82	97	85	18.32
PER-DDQN	80	98	81	18.57
Proposed method	93	96	97	17.96

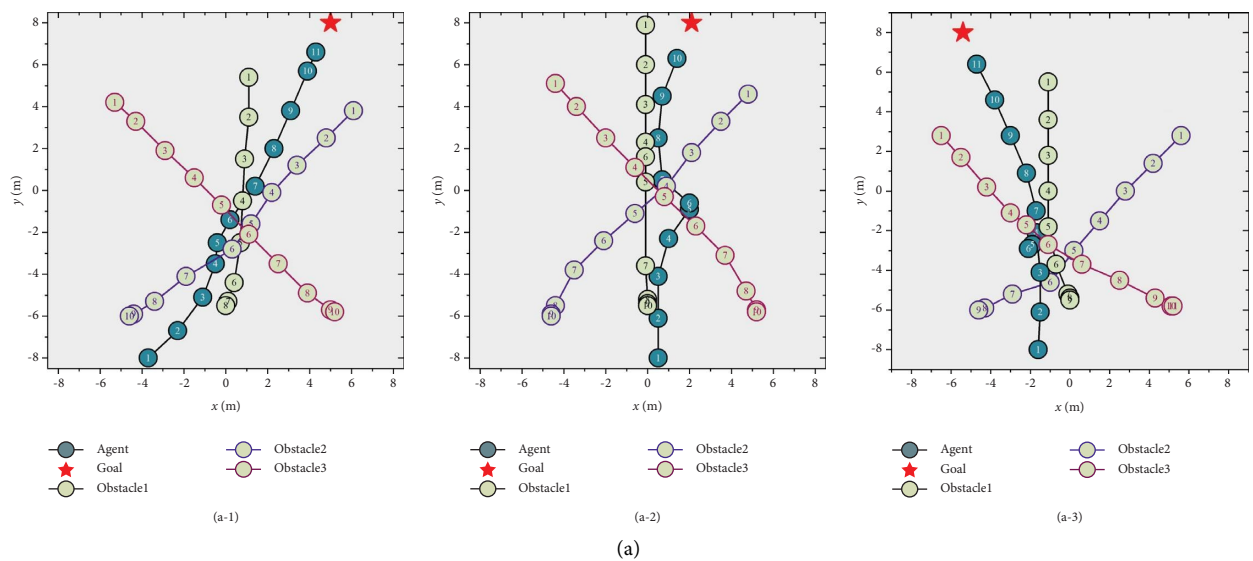


FIGURE 13: Continued.

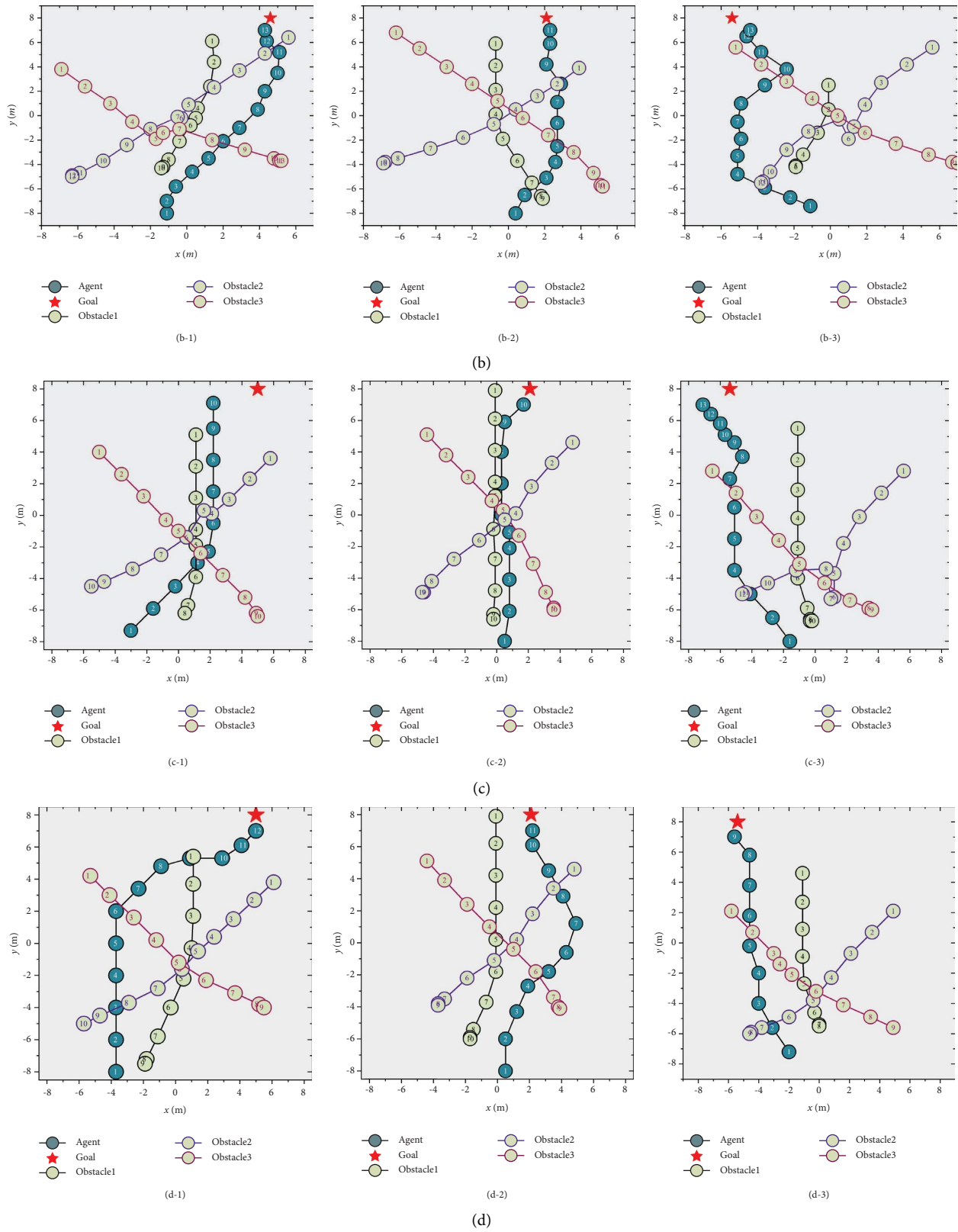


FIGURE 13: Path planning results of various path planning methods in the same environment: (a) A* + ORCA, (b) dueling DQN, (c) PER-DDQN, and (d) proposed method.

approaching the goal. The experimental results show that the method proposed in this paper is generally safe and performs well in various environments.

6. Conclusion

In this paper, a path planning algorithm for USVs in complex marine environments based on multiobjective reinforcement learning is proposed. To simulate complex ocean environment, a complex scene including dynamic obstacles and random goal is built. On this basis, two subtarget scenes with goal approaching and dynamic obstacle avoidance are established, respectively. The PER-DDQN algorithm is used to train the action decision of the agent in the two subtarget scenes. A multiobjective reinforcement learning architecture is designed to optimize the agent's policy integration method in path planning. The simulation results show that the proposed method achieves a higher path planning success rate and a shorter path length than the traditional path planning methods.

Although the proposed method realizes the decision making of the agents in the constructed scenes, the complexity of the scene is still insufficient. The computational efficiency and path planning success rate of the algorithm will be reduced in complex environments. Modelling more actual scenes and building more realistic training scenes can effectively improve the adaptability of the algorithm. In addition, the action space in the established model is discrete, which is somewhat different from the real world. Agents cannot output continuous action decisions in the scenario of discrete action strategy only. The assumption that the next time step after the action can reach the target position is also idealized, and inertial factors need to be taken into account to optimize the model. In future work, hostile ships with tracking capabilities will be added to the scene to train the model better. The dimension of the action space will be increased to enhance the USV's mobility. In addition, changing the scene from 2-dimensional space to 3-dimensional space is our follow-up research direction.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by Marine Defense Technology Innovation Center Innovation Fund (no. JJ-2020-719-02) and Knowledge Innovation Program of Wuhan-Basic Research.

References

- [1] J. Yuh, G. Marani, and D. R. Blidberg, "Applications of marine robotic vehicles," *Intelligent service robotics*, vol. 4, no. 4, pp. 221–231, 2011.

- [2] X. Liu, Y. Li, J. Zhang, J. Zheng, and C. Yang, "Self-adaptive dynamic obstacle avoidance and path planning for USV under complex maritime environment," *IEEE Access*, vol. 7, pp. 114945–114954, 2019.
- [3] Y. Liu and R. Bucknall, "Path planning algorithm for unmanned surface vehicle formations in a practical maritime environment," *Ocean Engineering*, vol. 97, pp. 126–144, 2015.
- [4] S. Guo, X. Zhang, Y. Du, Y. Zheng, and Z. Cao, "Path planning of coastal ships based on optimized DQN reward function," *Journal of Marine Science and Engineering*, vol. 9, no. 2, p. 210, 2021.
- [5] Y. Singh, S. Sharma, R. Sutton, D. Hatton, and A. Khan, "A constrained A* approach towards optimal path planning for an unmanned surface vehicle in a maritime environment containing dynamic obstacles and ocean currents," *Ocean Engineering*, vol. 169, pp. 187–201, 2018.
- [6] M. Y. Gao, B. B. Hu, and B. Liu, "Constrained path-planning control of unmanned surface vessels via ant-colony optimization," in *Proceedings of the 2021 40th Chinese Control Conference (CCC)*, pp. 4079–4084, IEEE, Shanghai, China, July 2021.
- [7] G. Xia, Z. Han, B. Zhao, and X. Wang, "Local path planning for unmanned surface vehicle collision avoidance based on modified quantum particle swarm optimization," *Complexity*, vol. 2020, Article ID 3095426, 15 pages, 2020.
- [8] Y. l Yao, X. f Liang, M. z Li et al., "Path planning method based on D* lite algorithm for unmanned surface vehicles in complex environments," *China Ocean Engineering*, vol. 35, no. 3, pp. 372–383, 2021.
- [9] X. Liang, P. Jiang, and H. Zhu, "Path planning for unmanned surface vehicle with dubins curve based on GA," in *Proceedings of the 2020 Chinese Automation Congress (CAC)*, pp. 5149–5154, IEEE, Shanghai, China, November 2020.
- [10] H. Zhai, W. Wang, and W. Zhang, "Path planning algorithms for USVs via deep reinforcement learning," in *Proceedings of the 2021 China Automation Congress (CAC)*, pp. 4281–4286, IEEE, Beijing, China, October 2021.
- [11] L. Xiong, Y. Kang, and P. Zhang, "Research on behavior decision-making system for unmanned vehicle[J]," *Automobile Technology*, vol. 515, no. 8, pp. 4–12, 2018.
- [12] H. Xu, N. Wang, H. Zhao, and Z. Zheng, "Deep reinforcement learning-based path planning of underactuated surface vessels," *Cyber-Physical Systems*, vol. 5, no. 1, pp. 1–17, 2019.
- [13] L. Tai, G. Paolo, and M. Liu, "Virtual-to-real deep reinforcement learning: continuous control of mobile robots for mapless navigation," in *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 31–36, IEEE, Vancouver, Canada, September 2017.
- [14] Y. F. Chen, M. Liu, and M. Everett, "Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning," in *Proceedings of the 2017 IEEE international conference on robotics and automation (ICRA)*, pp. 285–292, IEEE, Singapore, May 2017.
- [15] C. Chen, Y. Liu, and S. Kreiss, "Crowd-robot interaction: crowd-aware robot navigation with attention-based deep reinforcement learning," in *Proceedings of the 2019 International Conference on Robotics and Automation (ICRA)*, pp. 6015–6022, IEEE, Montreal, Canada, May 2019.
- [16] Z. Hu, Y. Zhang, Y. Xing, Y. Zhao, D. Cao, and C Lv, "Toward human-centered automated driving: a novel spatiotemporal vision transformer-enabled head tracker," *IEEE Vehicular Technology Magazine*, vol. 17, no. 4, pp. 57–64, 2022.

- [17] Z. Hu, Y. Xing, W. Gu, D. Cao, and C. Lv, "Driver anomaly quantification for intelligent vehicles: a contrastive learning approach with representation clustering," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 37–47, 2023.
- [18] K. Zhou, C. Yang, J. Liu, and Q. Xu, "Dynamic graph-based feature learning with few edges considering noisy samples for rotating machinery fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 10, pp. 10595–10604, 2022.
- [19] C. Yang, J. Liu, K. Zhou, X. Yuan, and M. F. Ge, "Transfer graph-driven rotating machinery diagnosis considering cross-domain relationship construction," *IEEE*, vol. 27, no. 6, pp. 5351–5360, 2022.
- [20] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers of Computer Science*, vol. 14, no. 2, pp. 241–258, 2020.
- [21] C. J. C. H. Watkins and P. Dayan, "Technical note: Q-learning," *Machine Learning*, vol. 8, no. 3/4, pp. 279–292, 1992.
- [22] V. Mnih, K. Kavukcuoglu, and D. Silver, "Playing atari with deep reinforcement learning," 2013, <https://arxiv.org/abs/1312.5602>.
- [23] T. Schaul, J. Quan, and I. Antonoglou, "Prioritized experience replay," 2015, <https://arxiv.org/abs/1511.05952>.
- [24] S. Zhang and R. S. Sutton, "A deeper look at experience replay," 2017, <https://arxiv.org/abs/1712.01275>.
- [25] F. Zhang, C. Gu, and F. Yang, "An improved algorithm of robot path planning in complex environment based on Double DQN," *Advances in Guidance, Navigation and Control*, pp. 303–313, Springer, Singapore, 2022.
- [26] Z. Huang, S. Liu, and G. Zhang, "The USV path planning of Dueling DQN algorithm based on tree sampling mechanism," in *Proceedings of the 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, pp. 971–976, Dalian, China, April 2022.
- [27] M. N. Alpdemir, "Tactical UAV path optimization under radar threat using deep reinforcement learning," *Neural Computing & Applications*, vol. 34, no. 7, pp. 5649–5664, 2022.
- [28] D. Lee, J. Kim, K. Cho, and Y. Sung, "Advanced double layered multi-agent Systems based on A3C in real-time path planning," *Electronics*, vol. 10, no. 22, p. 2762, 2021.
- [29] D. Xu, X. Liu, and Z. Huang, "The path-planning algorithm of unmanned ship based on DDPG[J]," *International Core Journal of Engineering*, vol. 8, no. 2, pp. 446–453, 2022.
- [30] H. Zhang, S. Bai, and X. Lan, "Hindsight trust region policy optimization," 2019, <https://arxiv.org/abs/1907.12439#:~:text=HTRPO%20leverages%20two%20main%20ideas,to%20select%20conductive%20hindsight%20goals>.
- [31] K. Chikhaoui, H. Ghazzai, and Y. Massoud, "PPO-based reinforcement learning for UAV navigation in urban environments," in *Proceedings of the 2022 IEEE 65th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 1–4, IEEE, Fukuoka, Japan, August 2022.
- [32] A. A. Khan, A. A. Laghari, T. R. Gadekallu et al., "A drone-based data management and optimization using metaheuristic algorithms and blockchain smart contracts in a secure fog environment," *Computers & Electrical Engineering*, vol. 102, Article ID 108234, 2022.
- [33] Y. Yang, W. Wang, and R. Xu, "AoI optimization for UAV-aided MEC networks under channel access attacks: a game theoretic viewpoint," in *Proceedings of the ICC 2022-IEEE International Conference on Communications*, pp. 1–6, IEEE, Seoul, Republic of Korea, May 2022.
- [34] Z. Zhu, C. Hu, C. Zhu, Y. Zhu, and Y. Sheng, "An improved dueling deep double-Q network based on prioritized experience replay for path planning of unmanned surface vehicles," *Journal of Marine Science and Engineering*, vol. 9, no. 11, p. 1267, 2021.
- [35] P. Zhai, Y. Zhang, and W. Shaobo, "Intelligent ship collision avoidance algorithm based on DDQN with prioritized experience replay under COLREGs," *Journal of Marine Science and Engineering*, vol. 10, no. 5, p. 585, 2022.
- [36] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

Research Article

A Framework and Algorithm for Human-Robot Collaboration Based on Multimodal Reinforcement Learning

Zeyuan Cai,^{1,2} Zhiquan Feng ,^{1,2} Liran Zhou,^{1,2} Changsheng Ai,³ Haiyan Shao,³ and Xiaohui Yang ^{1,4}

¹School of Information Science and Engineering, University of Jinan, Jinan 250022, China

²Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan 250022, China

³School of Mechanical Engineering, University of Jinan, Jinan 250022, China

⁴State Key Laboratory of High-end Server & Storage Technology, Jinan, China

Correspondence should be addressed to Zhiquan Feng; ise_fengzq@ujn.edu.cn

Received 12 May 2022; Revised 23 June 2022; Accepted 27 June 2022; Published 28 September 2022

Academic Editor: Zhongxu Hu

Copyright © 2022 Zeyuan Cai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Despite the emergence of various human-robot collaboration frameworks, most are not sufficiently flexible to adapt to users with different habits. In this article, a Multimodal Reinforcement Learning Human-Robot Collaboration (MRLC) framework is proposed. It integrates reinforcement learning into human-robot collaboration and continuously adapts to the user's habits in the process of collaboration with the user to achieve the effect of human-robot cointegration. With the user's multimodal features as states, the MRLC framework collects the user's speech through natural language processing and employs it to determine the reward of the actions made by the robot. Our experiments demonstrate that the MRLC framework can adapt to the user's habits after repeated learning and better understand the user's intention compared to traditional solutions.

1. Introduction

The limitations of “human-centered” or “robot-centered” robots have become increasingly essential with the advancement of the theories and applications of natural human-robot interaction. Regardless of the type of robots, understanding human intentions is indispensable for achieving complex human-robot collaboration processes [1]. A well-designed robot should dynamically adapt its behavior to different tasks and help humans more efficiently and respond to changes in the environment, humans, and own state in real time. In our daily life, we often communicate with others by expressing behavioral intentions [2–4] (such as gazes, gestures, and actions) to obtain a tacit understanding when collaborating with them. These intentions are frequently expressed through messages such as body language, voice, and mannerisms.

Perceiving the information mentioned above is not a complex problem for robots, and there have been many related studies making robots more effective in perceiving

people and environments [5]. However, this modal information obtained through sensors does not intuitively and accurately express real human intentions. Using this information rationally to acquire the exact human intentions is a popular research field at present.

2. Related Work

2.1. Human-Robot Collaboration and Intention Understanding. Many excellent algorithms have emerged in implementing human-robot collaboration based on intention understanding. In 2007, Suzuki et al. [6] estimated the intention of a paraplegic patient using the reaction forces on the floor while walking and standing. Their conclusions demonstrated the effectiveness of this approach in supporting the daily life of people with disabilities. In 2013, the Bayesian model was applied by Wang et al. [7] to estimate intentions on kinetic models generated in the process of human motion and recognition intention results were updated when additional motion data were obtained. In

2015, Ref. [8] employed manually applied forces and hip rotations to identify motion intention. Their conclusions verified that the method was effective in helping older people with turning assistance. In 2018, a hidden Markov model was used by Berg et al. [9] to recognize human motion, allowing robots to better adapt to human behavior and achieve human-robot synergy on production lines. Reference [10] utilized brain-computer interfaces (BCIs), visual interfaces, and remote robots to perform “emulated haptic shared control,” through which a remote proximity perception system was established to achieve human-robot collaboration and enable tetraplegic patients to interact with friends and the environment. Reference [11] performed excellent work in the opposite direction by proposing a method to interpret robot behavior as intention signals using natural language sentences, so as to better reveal robot behaviors and reduce misunderstandings caused by information asymmetry. Reference [12] proposed the proactive incremental learning (PIL) framework that learned the connection between human gestures and robot actions, which contributes to efficient human-robot interaction. In 2019, Ref. [13] proposed a cooperative fuzzy impedance control with embedded safety rules, a method that provides assistance and safety for human operators in heavy industrial scenarios. Reference [14] computed the probability distribution of intentions corresponding to each modality and then output these distributions through a Bayesian approach to combine independent opinion bases. The results suggested that this approach was better in accuracy and robustness than a unimodal-based classifier. In 2021, Ref. [15] achieved intention understanding using a single modality, gaze. Their study confirmed that gaze enabled more efficient human-computer collaboration in specific scenarios. In addition to conventional modalities (such as gesture, gaze, voice, and action), many studies exploited the information of less detectable modalities for intention understanding. For example, Ref. [16] designed a new approach for intention understanding of upper limb movements using mobile electroencephalography (EEG) via LSTM-RNN, which could provide early warning of impending danger to improve the safety of the system. Reference [17] constructed a projective recurrent neural network to estimate the joint angular intention of the user during motion using a Hill-based muscle model. Another interesting approach exploits the human’s preference to adapt the robot’s behavior based on the human’s feedback. Reference [18] discussed the necessity of user preferences in the design of robotic exoskeletons. Reference [19] proposed a path-based velocity planner, which uses the optimization method based on user pairwise preferences, can classify different paths and adjust the robot execution velocity more finely.

2.2. Reinforcement Learning. Reinforcement learning, as a field of artificial intelligence, has made significant progress since its introduction. An increasing number of human-robot collaborations use reinforcement learning to handle their problems. Reference [20] proposed the DQN algorithm in 2015. The core idea of the DQN algorithm is to use the

strong fitting property of neural networks to calculate the score of each action A in the input state S . It tackles the problem that the Q-Learning algorithm cannot handle scenarios with large state spaces. DQN takes two neural network structures of the same architecture: Q_{target} and Q_{eval} . The former intermittently updates the parameter θ^- , while the latter updates the parameter θ in real time. Additionally, DQN experience replay saves the experienced state-action pairs (s, a) , the corresponding reward, and next states in a memory bank, from which previous experiences randomly selected from the memory bank can be learned while the DQN is iterating. DDQN [21] is an improved version of DQN. Although DQN has been revealed to be effective in many applications, it still has some shortcomings. Since DQN uses the maximum operation to estimate the reward for the next state, DQN overestimates the Q value after several iterations. DQN strips the selection of the action and the evaluation of the action, estimates the best action in the Q_{eval} network parameters, and finds the score of the action in the Q_{target} network parameters. The authors put forward theoretical and experimental suggestions that DDQN effectively eliminates the drawbacks of DQN overestimation.

In 2019, Ref. [22] adopted the Soft Actor Critic reinforcement learning algorithm to build a robotic platform that the robot is capable of learning cooperative tasks with people in only 30 minutes without simulation training. Reference [23] proposed a multirobot path planning algorithm with deep q-learning combined with a convolutional neural network algorithm. Their simulation results revealed that the robot using this method had flexible and efficient motion performance in various environments compared to the traditional method. Reference [24] first applied deep reinforcement learning to the Urban Search and Rescue Team. They combined deep reinforcement learning with the robotic exploration of uncharted territory, which allowed the robot to explore the location environment autonomously. Their experiments demonstrated that the method could shrink victims faster than other methods. Inspired by Google Deep Mind, Ref. [25] formatted the collaborative human-robot assembly workflow as a chessboard. Specifically, the motion selection on the chessboard was used to simulate the human and robot decision-making in the human-robot collaborative assembly workflow. The self-training algorithm based on reinforcement learning was used for training, and the best strategy of collaborative human-robot work sequence was obtained without guidance or domain knowledge beyond the rules of the game, so as to improve efficiency. Reference [26] encoded the task and safety-related requirements into reinforcement learning and applied reinforcement learning to protect users in the process of human-robot cooperation. Reference [27] designed a human-centered collaborative system based on reinforcement learning that adopted unsupervised end-to-end learning to effectively tackle uncertainty in human behavior recognition and improve the behavioral decisions of the robot. In this way, the risks and benefits achieved by the robot after taking action were balanced. In 2020, Ref. [28] modeled the complex human-

computer interaction dynamics and proposed a model-based reinforcement learning variable impedance control, which minimizes the human force consumption. In 2021, Ref. [29] used reinforcement learning on dynamic task partitioning in assembly tasks with good results. Reference [30] presented an adaptive training method based on a deep Q-learning approach. The method treated robots and humans as agents in an interactive training environment for learning. Their algorithm addressed how to consider the dynamics of the time dimension and the stochasticity in the program sequence when the collaborative assembly was an industrial task.

Reinforcement learning has demonstrated its powerful role in robotics by enabling robots to learn independently [31]. In summary, human-robot collaboration cannot be separated from the process of intention understanding. The fusion of multiple modalities has been discovered to have higher robustness in the correctness of intention understanding. Most human-robot collaboration frameworks based on intention understanding have explored an intention understanding paradigm working for all users to improve the robustness of the algorithm and the accuracy of intention understanding. However, it is tough to find a suitable paradigm for all users, as each user has different habits in expressing their intentions due to the variability among individuals. It is a problem that needs to be solved to make the robots work efficiently with various users.

To this end, a multimodal human-robot collaboration (MRLC) framework based on reinforcement learning is proposed in this article. MRLC is divided into two parts. The first part is the intention understanding process based on the deep reinforcement learning algorithm (DDQN) [21], which adapts to the habits of individuals through iterative training and forms habit rules for each user. Hence, our collaboration framework can eliminate the issues of inconsistent collaboration when facing users with different habits in traditional human-robot collaboration. The second part is the task assignment process. Through the first part of intention understanding, the robot understands “what the user wants to do” and then MRLC assigns tasks to the robot to collaborate with the user.

This article has three main contributions:

- (1) The MRLC human-robot collaboration framework is proposed. Reinforcement learning is adopted for intention recognition to make the human-robot collaboration framework more robust to different users.
- (2) The MRLC framework uses a reward function incorporating natural language processing, which maintains the user experience during robot’s learning process.
- (3) The MRLC human-robot collaboration framework is successfully applied to a scenario where humans and robots work together to build a Jenga tower.

A human-robot collaboration scenario is designed for building the Jenga tower to verify the feasibility of our

proposed collaboration framework. In this scenario, the robot needs to consider the state of the Jenga tower and user’s intention to achieve dynamic collaboration among the robot, Jenga tower, and user.

3. Materials and Methods

3.1. MRLC Structure. Existing human-robot collaboration frameworks mainly use a unified paradigm to observe user’s characteristics to achieve intention understanding and human-robot collaboration. In contrast, the MRLC framework has the following features. (1) MRLC is capable of adapting to each user’s behavioral habits rather than applying a uniform paradigm to every user. (2) It enables a three-way interaction between robot and user, user and environment, and robot and environment. The structure of MRLC is illustrated in Figure 1. The main goal of MRLC is to allow the robot to learn user’s behavioral habits, to recognize user’s intentions, and to assign tasks to the user’s following different intentions. It guarantees that the robot can perform human-robot collaboration tasks dynamically and safely.

The MRLC framework is divided into two main modules: multimodal intention understanding module and task assignment module. First, the robot collects the characteristics of user’s three modalities and learns user’s behavioral habits through reinforcement learning to obtain user’s behavioral intention. After user’s intention is obtained, the robot enters the task assignment phase, in which the robot’s action sequence is specified based on user’s behavior. Then, the robot starts to interact with the environment and the user.

3.2. Multimodal Intention Understanding Based on Reinforcement Learning. The MRLC framework contains a novel multimodal reinforcement learning intention understanding algorithm. The core idea is to learn users’ behavioral habits through deep reinforcement learning in iterative iterations, so as to eliminate errors induced by differences in behavioral habits of different users and to achieve a more robust intention understanding.

Figure 2 illustrates the architecture of the multimodal reinforcement learning intention understanding algorithm. It is divided into three stages: (1) extraction of user multimodal features. The data obtained from the sensors first go through three subclassifiers to obtain the classification results of m_1 , m_2 , and m_3 . The user modal data are finally converted into a 3D vector $\mathbf{s} = [m_1, m_2, m_3]$; (2) the extracted user features are used as state inputs to fit the scores under each intention outcome; (3) calculation of the optimal operation corresponding to the user intention according to the optimization objective by equation (1), followed by the analysis of the user’s linguistic feedback using NLP to obtain the user satisfaction S_{θ} , which is learned iteratively as part of the reward:

$$i = \max Q(\mathbf{s}, \mathbf{I}; \theta), \quad (1)$$

where \mathbf{s} represents the user’s features, i represents the best intention at moment, \mathbf{I} represents intentions spaces, Q represents the value of each intention calculated using the q_val neural network, and θ represents the parameter of the q_val neural network:

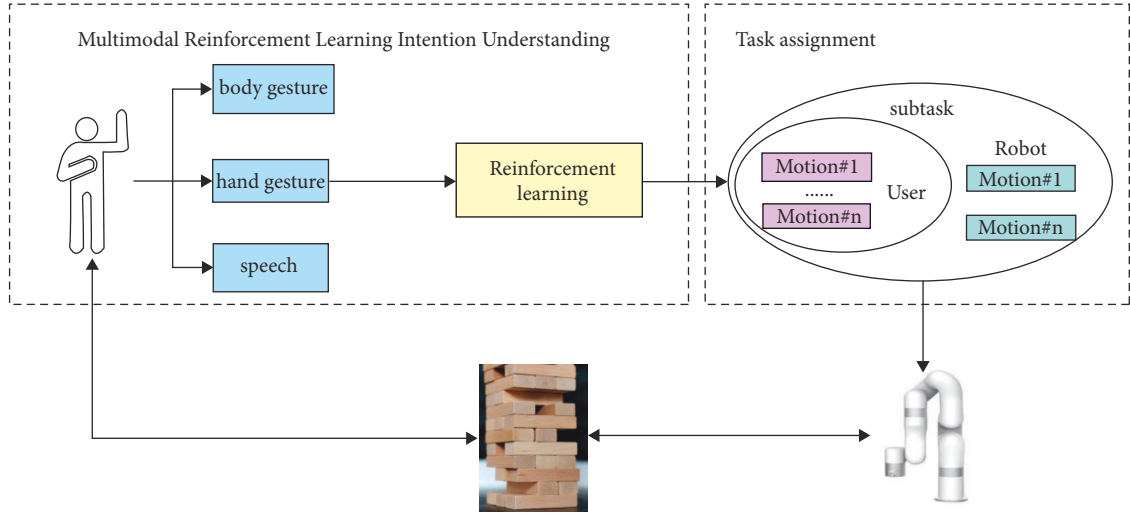
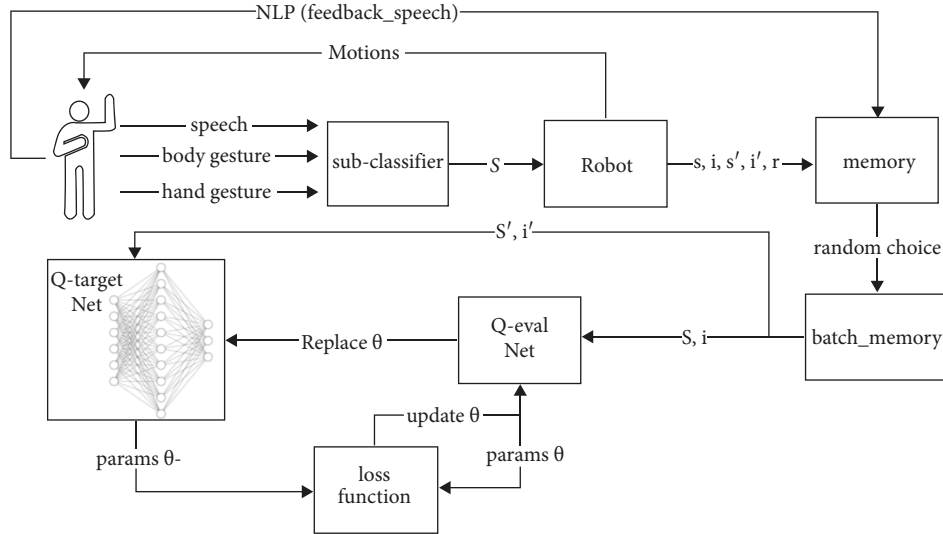


FIGURE 1: The structure of MRLC.

FIGURE 2: Multimodal reinforcement learning intention understanding algorithm architecture. s represents the user's features, s' represents the user's features next time, i represents the result of intention understanding, and i' represents the result of intention understanding next time.

3.3. User Multimodal Feature Extraction. The user's behavioral characteristics are defined as autonomous, natural movements (such as gestures and speeches) made during human-robot collaboration.

Three sensors are arranged to implement the user's input in three modalities: speech, body gesture, and hand gesture. In our work, a stable and efficient data processing method is selected to process the data obtained from the sensors. Regarding speech modality, the user's speeches are converted into text and classified into seven categories by combining keyword recognition. Another interesting approach is the use of pointing to achieve natural human-computer interaction. There has been a lot of outstanding work [32, 33] on their methods to detect user pointing accurately. However, considering the complexity of the system, we do not apply these results in our system. The category numbers

corresponding to the speech keywords are listed in Table 1. Concerning body gesture modality, KinectGesture in KinectV2 is adopted to implement the recognition of four types of static user body gestures. With respect to hand gesture modalities, efficientnetV2 is employed to implement five types of hand gesture recognition. All the body gestures and hand gestures that can be detected are exhibited in Figures 3 and 4.

3.4. Reward Functions Based on Natural Language Processing. In the process of human-robot collaboration, it is tricky to determine an appropriate reward function while ensuring the user experience. The robot needs to know whether its behavior is correct during the learning process. Telling the robot whether the behavior is correct every time by manual input would cut off the user experience.

TABLE 1: User’s keywords and category numbers.

Keywords	Category numbers
Stop	1
Take a block	2
Put the block on	3
Hand me the block	4
Take the block	5
Put the block aside	6
There is your block	7

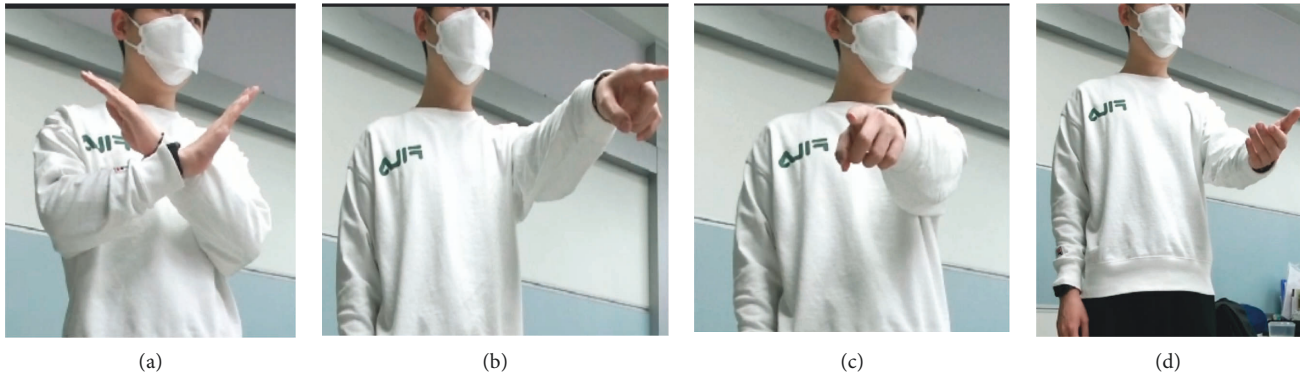


FIGURE 3: User’s body gesture. (a) Cross arms; (b) point to the unplaced pile of blocks; (c) point to the built Jenga tower; (d) raise your hand in a small increment.

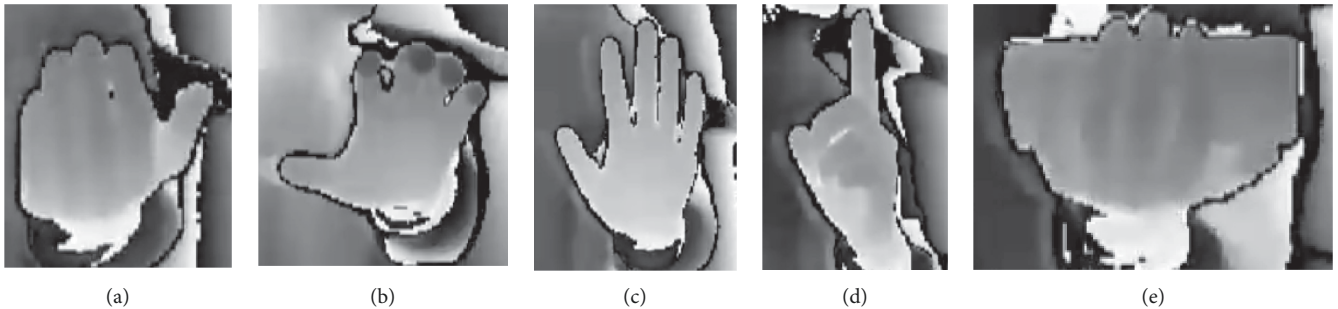


FIGURE 4: User’s hand gestures. (a) Fingers bent slightly upward; (b) fingers bent slightly downward; (c) open palm; (d) index finger up; (e) pick up a block.

Therefore, natural language processing techniques are utilized to collect users’ evaluations: NLP (feedback_speeches). With the snownlp module, speech emotion analysis is performed on the speech collected by the microphone. The result of NLP (feedback_speeches) is between 0 and 1 and is judged as positive feedback when the result is higher than 0.5. The intention is understood correctly when the user’s evaluation is positive (such as “good job”). Notably, if the user does not give any feedback, we believe the user’s tacit approval of the behavior and regard it as positive feedback:

$$R = \begin{cases} 10, & \text{NLP}(\text{feedback_speeches}) > 0.5 \text{ or nonexistent,} \\ 10, & \text{NLP}(\text{feedback_speeches}) < 0.5. \end{cases} \quad (2)$$

3.5. Multimodal Reinforcement Learning Intention Understanding Algorithm Network Structure. The multimodal reinforcement learning intention understanding algorithm proposed in this article has two neural networks with the same structure: q_target and q_eval [21], both of which consist of two fully connected layers $l1$ and $l2$. Among them, $l1$ consists of 50 neurons. The multimodal reinforcement learning intention understanding algorithm also adopts a memory bank to store the previously learned results and implement the offline learning. The input to q_eval is the user features s , which are also the results of the three subclassifiers. q_eval first fits the user features s using random weights to derive a score for each state. The intention with the highest score is selected as the best result for

output. After the user gives feedback, the reward value R is derived according to the reward function (2). Then, the sum of current rewards and expected future rewards y' is calculated by the following equation:

$$y' = R + \gamma * \max_i Q(s', \max_i Q(s', I; \theta)); \theta^-, \quad (3)$$

where γ represents the decay factor of future reward, θ represents the parameter of q_eval net, θ^- represents the parameter of q_target net, R represents the current reward, s' designates the multimodal input at the next intention understanding, and I represents the intentions spaces.

Since the update frequencies of q_target and q_eval are different, the following loss function can be obtained by using Temporal-Difference (TD).

$$Loss = (y' - Q(s, i; \theta))^2, \quad (4)$$

where s denotes the multimodal input and i represents the result of intention understanding.

3.6. Task Assignment. In the previous section, the user's intention has been derived based on the multimodal reinforcement learning intention understanding algorithm. In this section, task assignment is performed based on the intention.

The MRLC framework uses a top-down, progressively refined dynamic task assignment approach as illustrated in Figure 5. Specifically, a database $M(\text{intention}, \text{task})$ of intentions and subtasks, and a database $M(\text{subtask}, \text{motion})$ of subtasks and actions are constructed. The final task is progressively refined to all pending action sequences Motion.

A reasonable task assignment module can dynamically assign tasks to the robot based on the user's behavior instead of rigidly specifying the tasks that the robot needs to be responsible for. Under the concept of sets in mathematics, all tasks are considered a full set Motion. The tasks that the user has completed are a subset $\text{Motion}_{\text{user}}$. Then, the tasks that the robot needs to be responsible for are the complement of $\text{Motion}_{\text{user}}$ where $\text{Motion}_{\text{robot}} = \text{Motion} - \text{Motion}_{\text{user}}$. This approach allows the MRLC framework to achieve dynamic task assignments to further increase the flexibility of collaboration. Additionally, the MRLC framework can be easily applied to other collaboration scenarios by modifying the two databases.

For example, if $\text{Motion}_{\text{user}}$ is {"user picks up a block"} and Motion is {"user picks up a block", "robot moves towards user's hand", "robot grabs the block in user's hand"}, then $\text{Motion}_{\text{robot}}$ is {"robot moves towards user's hand", "robot grabs the block in user's hand"}.

3.7. Algorithm Analysis. Based on the above research ideas and the MRLC architecture diagram in Figure 1, a specific description of the MRLC architecture algorithm1 is presented as follows.

The MRLC framework aims to eliminate the bias in collaboration effectiveness caused by the variability of individual user habits. Multimodal reinforcement learning intention understanding methods are employed to learn

each user's habits and thus weaken the impact of individual differences. One of the crucial metrics to evaluate the effectiveness of the MRLC framework is the correct rate of intention understanding, which is the core of the MRLC framework's ability to adapt to different user habits.

When a new user tries to collaborate with the robot for the first time in our human-robot collaboration scenario, the multimodal reinforcement learning intention understanding algorithm is set up to first sense the three modal information of the user and use it as input to predict the user's intention and perform task assignment; besides, the parameters of the algorithm are adjusted relying on the feedback given by the user; in this way, the issue of how to make the robot learn the user's habits is theoretically overcome. With an increase in the number of learning times, the multimodal reinforcement learning intention understanding algorithm gradually converges. The intention understanding becomes more effective, suggesting that the MRLC framework learns the user's habits.

The MRLC framework implements the perception of different modal data through an efficient subclassifier, instead of feeding the collected basic information directly into the deep reinforcement learning neural network to ensure real-time collaboration. The multimodal reinforcement learning intention understanding algorithm only needs to process a three-dimensional matrix representing multimodal information, which contributes to a significant decrease in the time complexity and ensure the real-time performance of the algorithm.

The MRLC framework theoretically addresses the issue presented in this article: how can robots still maintain efficient collaboration facing users with different habits?

4. Experimental Results and Analysis

4.1. Experimental Scenes. Our human-robot collaboration scenario is shown in Figure 6, which consists of the Xarm 7-axis mechanical arm, mechanical gripper AG-95, two RGB cameras, and an RGB-D camera, where two RGB cameras are used to detect the status of the Jenga tower. The computer's CPU, GPU, and RAM are I7-10875H, RTX2060, and 16G, respectively. Furthermore, microphones and an RGB-D camera are employed to capture the user's speeches, body gestures, and hand gestures.

4.2. Experimental Scenes. Ten experimenters, including six males and four females, with an average age of 25 years were invited to participate in the experiment. They had never been exposed to similar human-robot collaboration scenarios before. Before the experiment, we informed the experimenters about the modalities that the robot could perceive some specific modal data, such as specific hand gestures and keywords.

The task of human-robot collaboration is to establish Jenga tower rather than playing Jenga game. Six categories of intentions are set. Table 2 lists all intentions and the corresponding numbers.

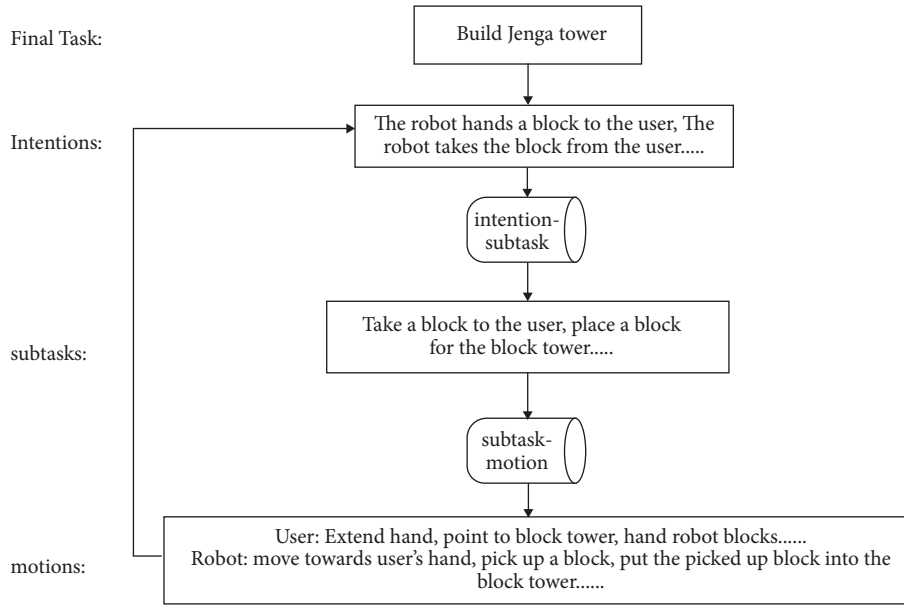


FIGURE 5: Task assignment process.

```

Input: User_speeches, User_body gestures, User_hand gestures, final_task, M(I, subtask), M(subtask, motion)
Initialize: NLP, sub_classifier, memory M, episode←0, load  $\theta$ , Sub_classifier (User_speeches, User_body gestures, User_hand gestures), replace_iter
Output: Motionrobot.
While not finishing final_task do:
  s ← Sub_classifiers
  With probability  $\epsilon$  to select a random intention  $i$ 
  Otherwise use equation (1) to calculate  $i$  subtask ←  $M(i, subtask)$ 
  Motion ←  $M(subtask, motion)$ 
  Motionrobot ← Motion – Motionuser  $r$  ← NLP (feedback_speech)
  //s' is the next behavior feature of User after robot executes Motionrobot
  s' ← Sub_classifiers after Robote executes (Motionrobot)
  Calculate Reward  $r_t$  according to equation (2)
   $M \leftarrow (s, i, r, s')$  batch_memory ← random choice (M)
  If s means the end of collaboration:
    y' ← r
  Else:
    Use equation (3) to calculate y'
    Use equation (4) to calculate loss
    Minimize loss
  If (episode > replace_iter):
     $\theta^- \leftarrow \theta$ 
End
  
```

ALGORITHM 1: MRLC Multimodal Reinforcement Learning Cooperation

A complete collaborative process is recorded, as exhibited in Figure 7. The robot and the user are in a face-to-face position, and the unplaced Jenga blocks are stacked on the side of the robot where the robot can easily clip them.

4.3. Experiment Procedure

4.3.1. The Differences in User Habits When Expressing Intentions. It is necessary to demonstrate the differences in expression habits of different users when expressing the same

intention. Thus, a questionnaire was distributed to the experimenters before the formal experiment to verify the specific modal categories used by each experimenter in expressing the same intention, in which we marked the specific body gestures, hand gestures, and keywords in Section 3.2.1. The table of the questionnaire is provided in Table 3.

4.3.2. The Success Rate of Human-Robot Collaboration. In this section, the experiment is performed on the relationship between the success rate of human-robot

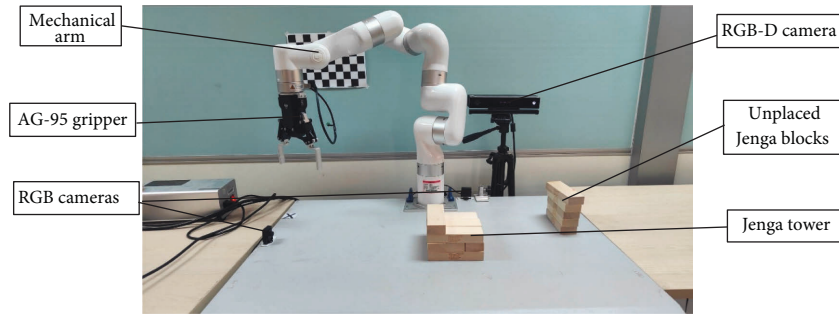


FIGURE 6: Schematic diagram of the experimental scene.

TABLE 2: Intentions and numbers.

Intentions	Numbers
The robot stops immediately	1
Robot takes a block and gives it to the user	2
Robot actively picks up a block	3
The robot takes the blocks from the user	4
Putting aside the blocks that the robot clips up	5
Put the blocks that the robot clips up onto the Jenga tower	6

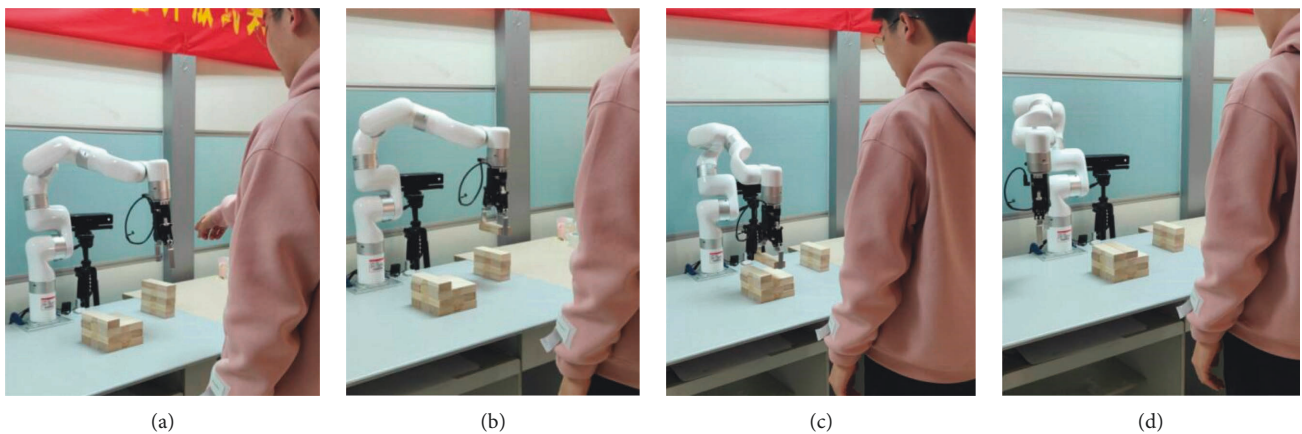


FIGURE 7: A typical human-robot collaboration process. (a) The user makes an act of pointing to an unplaced block and speaks the speech keyword “put the block on”; (b) the multimodal reinforcement learning intention understanding algorithm derives the user’s intention: the robot actively picks up the blocks; (c) the user gives the speech evaluation: “That is it” NLP evaluates this feedback to be positive; (d) the robot is reset, and the reinforcement learning learns the user’s habit and the round of collaboration ends.

TABLE 3: The questionnaire.

Intentions	Body gestures that can express the intention (if none, fill in 0)	Hand gestures that can express that intention (if none, fill in 0)	Keywords that can express the intention (if none, fill in 0)
The robot stops immediately			
The robot takes a block and gives it to the user			
The robot actively picks up a block			
The robot takes the block from the user			
Put aside the block that the robot clips up			
Put the blocks that the robot clips up onto the Jenga tower			

collaboration and learning times, as well as the success rate of human-robot collaboration in the MRLC framework for new users with different habits.

To explore the relationship between the success rate of human-robot collaboration, ten people are divided into five groups: A1, A2, A3, A4, and A5; we limit the number of times the robot learns to 200, 400, 600, 800, and 1000 for the A1, A2, A3, A4, and A5. At the end of the learning phase, the experimenter performs 100 tests using the learned parameters θ . The user feedback is positive as a successful human-robot collaboration. Moreover, the success rate of human-robot collaboration for each group is recorded.

Equation (5) is used to calculate the success rate of human-robot collaboration:

$$\text{Acc} = \frac{\text{count}_T}{100}, \quad (5)$$

where count_T indicates the number of successful human-robot collaborations during the test.

To explore the success rate of human-robot collaboration, ten people are divided into groups B and C, with two people in group B and eight people in group C. The experimenters in group B are divided into B1 and B2, with one experimenter in each group; the experimenters in group C are divided into C1, C2, C3, and C4, with two experimenters in each group. First, the robot learns the habits of the experimenters in group B 800 times, and then, we switch the users to the experimenters in group C. The experimenters in group C1 perform 100 tests without learning; the experimenters in group C2 perform 100 tests after 400 times of learning; the experimenters in group C3 perform 100 tests after 600 times of learning; and the experimenters in group C4 perform 100 tests after 800 times of learning. This process is adopted to simulate the robot's performance in an actual situation when it is confronted with a new user with different habits.

4.3.3. Comparison of Multimodal Reinforcement Learning Intention Understanding Algorithms with Naive Bayesian Intention Understanding Algorithms. The core of the MRLC framework lies in the multimodal reinforcement learning intentional understanding algorithm, by which the MRLC framework can learn the user's habits. Therefore, the multimodal reinforcement learning comprehension algorithm is compared with the classical traditional intention understanding algorithm (the naive Bayes intention understanding algorithm) in this section. Bayesian decision formula (6) is employed to calculate the probability of each intention when the user expresses the intention, and the intention with the highest probability is taken as the final result:

$$w_j|X = \frac{P(\mathbf{X}|w_j)P(w_j)}{\sum_{n=1}^N P(\mathbf{X}|w_n)P(w_n)} = \frac{\prod_{k=1}^M P(x_k|w_j)P(w_j)}{\sum_{n=1}^N \prod_{k=1}^M P(x_k|w_n)P(w_n)}, \quad (6)$$

where w_j represents the results of intention understanding, and \mathbf{X} refers to the user input multimodal data matrix, and x_k represents the k th component of \mathbf{X} .

Meanwhile, the MRLC framework was used to let the robot learn the user's habit from scratch. Ten experimenters were divided into five groups: D1, D2, D3, D4, and D5, each group of two. One experimenter User #1 was randomly selected from each group to collaborate with the robot using the multimodal reinforcement learning intention understanding algorithm. Starting from a learning count of 300, the robot paused learning every 200 times and tested the accuracy of the multimodal reinforcement learning intention understanding algorithm and the naive Bayes intention understanding algorithm 100 times, for a total of three times. After the 700 learning of User #1 was finished, the robot was replaced by another experimenter from the same group, User #2, to collaborate with the robot. This process stimulated the realization of a real scenario where the robot is confronted with a new user with different habits. The second experimenter was the same as the first one, starting from the 300th learning, pausing learning every 200 times, and performing 100 tests, for a total of three tests.

4.3.4. Evaluating MRLC with Human Factors. The degree of success of human-robot collaboration depends on the joint consideration of the robot factor (RF) and the human factor (HF) [34]. The experiment mentioned above is an evaluation of the robot factor. Therefore, this section uses four human factors indicators to evaluate the MRLC qualitatively. The four indicators are trust, anxiety, safety perception, and fatigue [34]. The score interval of each indicator is [1–10], and the lower score means the worse performance under the indicator, 0 means very poor, and 10 means very good. We collected the subjective feelings of all experimenters during the experiment through a questionnaire. It should be noted that during this experiment, we collected the subjective feelings of experimenters after MRLC fully learned the experimenter's habits.

4.4. Experiment Results

4.4.1. Differences in User Habits When Expressing Intentions. The data collected from the questionnaires in Section 4.3.1 were organized in the heat map (Figure 8), with a total of 10 questionnaires received. The maximum number of identical expressions in the same modality was recorded under the same intention. For example, in intention #1, if 7 questionnaires were selected to use one outcome under the modality of body gestures, the value for that position was 7.

Figure 8 reveals that users can reach a consensus on expression habits for specific intentions. For instance, the number of experimenters who choose to use the same expression reaches 8–9 in intention #1, implying that most users tend to express intentions in the same expression.

However, the expression habits vary significantly among users for some intentions, such as intention #6 and intention #3. Since traditional intention understanding algorithms, such as SVM and naive Bayes, aim at demonstrating correlations between modal data and intentions, this phenomenon can tremendously degrade the performance of traditional intention understanding algorithms.

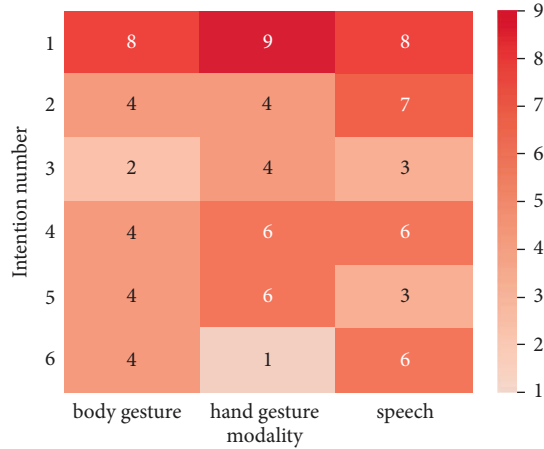


FIGURE 8: Maximum number of identical expressions with the same intention in the same modality.

4.4.2. Human-Robot Collaboration Success Rate. The data during the experiment are recorded in Table 4, involving the number of failed MRLC algorithm human-robot collaboration trials for individual experimenter of each group. Furthermore, each group's success rate of human-robot collaboration is plotted in Figure 9.

As observed in Table 4, the failure number of human-robot collaboration gradually decreases as learning times increase when learning times are between 400 and 850. The result reflects that the robot is gradually learning the user User #2's habits. Meanwhile, the intention understanding is gradually becoming more accurate. However, the failure number of human-robot collaboration increases when learning times rise to 1000. The principal reason is that the overfitting of the MRLC algorithm leads to a decrease in the effectiveness of human-robot collaboration.

A similar conclusion can be drawn from Figure 9. The success rate of human-robot collaboration in groups A1, A2, A3, and A4 rises and finally reaches 92%. The human-robot collaboration system achieves acceptable levels.

Table 5 presents the change in the failure number of human-robot collaboration with learning after changing experimenters with different C groups of habits.

Table 5 and Figure 10 demonstrate a significant decrease in the success rate of human-robot collaboration when the robot faces a user with different habits. In other words, the robot cannot effectively understand the user's intention when facing a new user, which is consistent with our expectation. The success rate of human-robot collaboration increases as learning times increase in the interval of 400 to 800 times of learning, which suggesting that the robot is continuously adapting to new user's habits. The robot reaches a success rate of 93% after 800 times of learning.

As revealed in the previous experiment, the success rate of human-robot collaboration decreases to a specific level when the number of learning times reaches 1000, which is ascribed to the overfitting of MRLC.

This experiment implies that the MRLC framework is malleable and achieves a success rate of more than 90% after hundreds of learning times for a new user's habit.

TABLE 4: Number of failures for different experimenters' human-robot collaboration results.

Times of learning/group	User #1	User #2
400/A1	30	21
550/A2	13	16
700/A3	12	13
850/A4	9	7
1000/A5	20	17

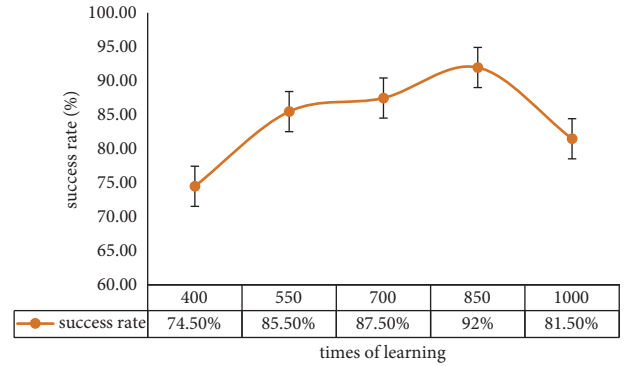


FIGURE 9: Success rate with the change in learning times.

TABLE 5: The failure numbers of human-robot collaboration with the change in learning times when facing new users.

Times of learning/Group	User #1	User #2
0/C1	75.5	78
400/C2	27	21.5
600/C3	21	19
800/C4	8.5	6
1000/C5	17	11.5

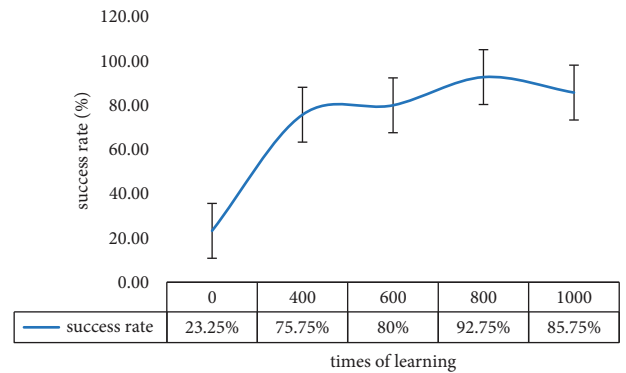


FIGURE 10: Variation in the human-robot collaboration success rate with the change in learning times when facing new users.

4.4.3. Comparison of Multimodal Reinforcement Learning Intention Understanding Algorithms with Naive Bayesian Intention Understanding Algorithms. As illustrated in Figure 11, the multimodal reinforcement learning intention understanding algorithm achieves 63.2% accuracy of correct intention understanding after 300 times of learning, which is 9.2% lower than the naive Bayes algorithm. However, after 500 times of learning, the multimodal reinforcement learning intention understanding algorithm reaches 84.2% accuracy of

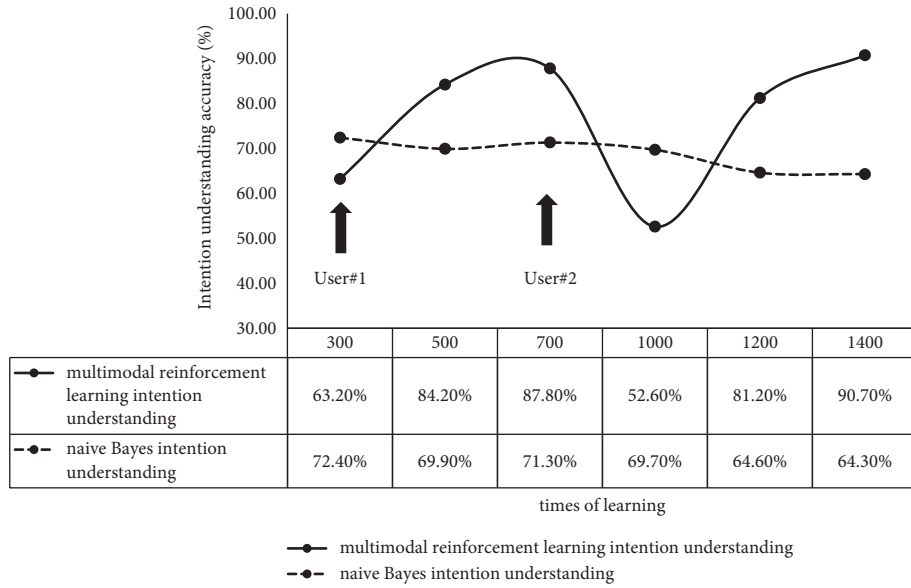


FIGURE 11: The comparison of accuracy variation of two algorithms with the change in learning times.

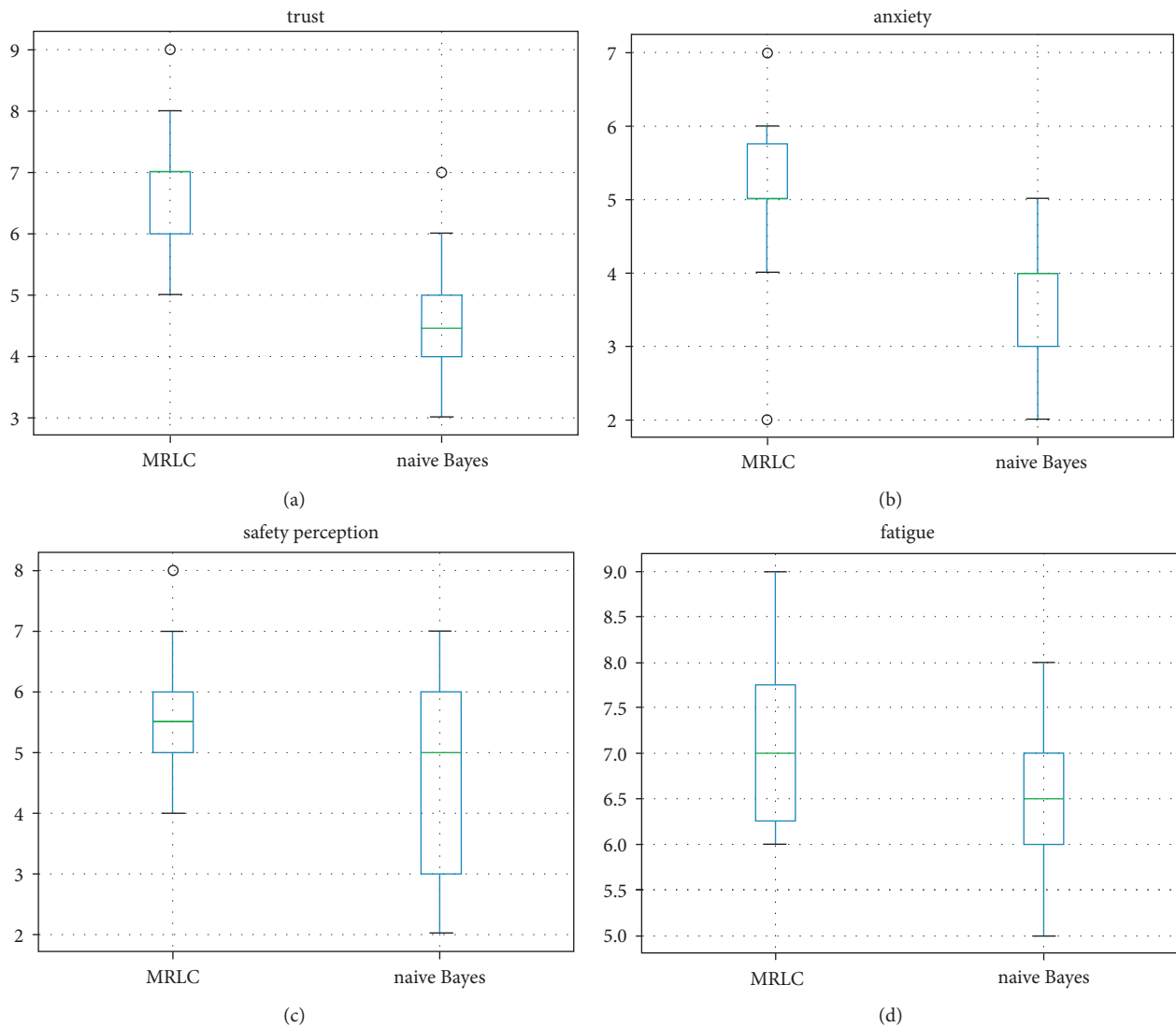


FIGURE 12: Human factors questionnaire results. (a) Trust; (b) anxiety; (c) safety perception; (d) fatigue.

intention recognition compared to the other method, with 14.3% performance improvement. The performance difference reaches 16.5% after 700 times of learning.

After changing experimenter User #2, the accuracy of the multimodal reinforcement learning intention understanding algorithm decreases to 52.6% which is lower than 69.7% accuracy of the naive Bayes algorithm. This is induced by the difference in habits between the experimenters. With the increasing learning times, 90.7% accuracy is reached after 700 times of learning, which is significantly higher than 64.3% accuracy of the naive Bayes algorithm.

This experiment demonstrates that the multimodal reinforcement learning intention understanding algorithm eliminates the inconsistent performance of traditional intention understanding algorithms for different users by providing accuracy over 90% after enough times of learning.

4.4.4. Evaluating MRLC with Human Factors. Ten questionnaires are collected in this article. Figure 12 illustrates the results of the assessment of human factors indicators for both frameworks. In the indicator of “trust,” MRLC performed significantly better, and most of the experimenters felt that the MRLC collaboration framework brought them more trust. There is not much difference between the two on “anxiety,” with MRLC performing slightly better and a significant portion of the experimenters scoring within 2 points. Neither a collaborative framework was effective in reducing user anxiety with the robot. Similar conclusions are found for the two indicators of safety perception and fatigue. Considering the differences between the experimenters and the resulting errors, there is no significant difference between the two algorithms.

5. Conclusion

Most of the traditional human-robot collaboration frameworks specify the process of human-robot collaboration, which corresponds to the user’s instructions and the robot’s actions.

The MRLC framework has the following advantages over traditional human-robot collaboration frameworks: (1) greater flexibility and higher adaptability. The multimodal reinforcement learning intention understanding algorithm achieves intention understanding in human-robot collaboration and thus solves the problem that the efficiency of traditional human-robot collaboration frameworks decreases when facing different user habits. (2) Stronger reusability. The MRLC framework can be easily applied to other human-robot collaboration scenarios. With the addition of reinforcement learning algorithms, users do not need to make an extra effort on editing the rules of human-robot collaboration. Concurrently, users can directly modify the database between layers to achieve dynamic human-robot task assignments owing to the hierarchical design of the task assignment module.

The experiments suggest that the success rate of collaboration in the MRLC framework reaches more than 90% after many times of learning, which improves over 10% compared with the traditional algorithm.

In our experiments, users are required to learn more than 800 times to achieve an excellent synergistic effect. Since the MRLC framework is based on deep reinforcement learning, it also inherits the shortcomings of deep reinforcement learning algorithms, such as slow convergence.

Therefore, the issues of slow learning speed and slow convergence of the MRLC framework should be overcome in the following research.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This article was supported by the Independent Innovation Team Project of Jinan City (Grant no. 2019GXRC013) and the Shandong Provincial Natural Science Foundation (Grant no. ZR2020LZH004).

References

- [1] D. Nocolis, A. M. Zanchettin, and P. Rocco, “Human Intention Estimation Based on Neural Networks for Enhanced Collaboration with Robots,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1326–1333, Madrid, Spain, January 2018.
- [2] S. J. Blakemore and J. Decety, “From the perception of action to the understanding of intention,” *Nature Reviews Neuroscience*, vol. 2, no. 8, pp. 561–567, 2001.
- [3] Z. Hu, Y. Zhang, Y. Xing, and Z. Yifan, “Toward Human-Centered Automated Driving: A Novel Spatial-Temporal Vision Transformer-Enabled Head Tracker,” *IEEE Vehicular Technology Magazine*, pp. 2–9, 2022.
- [4] Z. Hu, Y. Xing, W. Gu, D. Cao, and C. Lv, “Driver anomaly quantification for intelligent vehicles: a contrastive learning approach with representation clustering,” *IEEE Transactions on Intelligent Vehicles*, p. 1, 2022.
- [5] T. Li, X. Sun, X. Shu et al., “Robot grasping system and grasp stability prediction based on flexible tactile sensor array,” *Machines*, vol. 9, no. 6, p. 119, 2021.
- [6] K. Suzuki, G. Mito, H. Kawamoto, Y. Hasegawa, and Y. Sankai, “Intention-based walking support for paraplegia patients with Robot Suit HAL,” *Advanced Robotics*, vol. 21, no. 12, pp. 1441–1469, 2007.
- [7] Z. Wang, K. Mülling, M. P. Deisenroth et al., “Probabilistic movement modeling for intention inference in human-robot interaction,” *The International Journal of Robotics Research*, vol. 32, no. 7, pp. 841–858, 2013.
- [8] Y. H. Hsieh, K. Y. Young, and C. H. Ko, “Effective maneuver for passive robot walking helper based on user intention,” *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6404–6416, 2015.
- [9] J. Berg, T. Reckordt, C. Richter, and G. Reinhart, “Action recognition in assembly for human-robot-cooperation using hidden Markov models,” *Procedia CIRP*, vol. 76, pp. 205–210, 2018.

- [10] M. P. Pacaux-Lemoine, L. Habib, and T. Carlson, "Human-robot Cooperation through Brain-Computer Interaction and Emulated Haptic Supports," in *Proceedings of the IEEE International Conference on Industrial Technology (ICIT)*, pp. 1973–1978, Lyon, France, April 2018.
- [11] Z. Gong and Y. Zhang, "Behavior Explanation as Intention Signaling in Human-Robot Teaming," in *Proceedings of the 2018 27th IEEE International Symposium On Robot And Human Interactive Communication (RO-MAN)*, pp. 1005–1011, Nanjing, China, August 2018.
- [12] D. Shukla, Ö Erkent, and J. Piater, "Learning semantics of gestural instructions for human-robot collaboration," *Frontiers in Neurorobotics*, vol. 12, p. 7, 2018.
- [13] L. Roveda, S. Haghshenas, M. Caimmi, N. Pedrocchi, and L. Molinari Tosatti, "Assisting operators in heavy industrial tasks: on the design of an optimized cooperative impedance fuzzy-controller with embedded safety rules," *Frontiers in Robotics and AI*, vol. 6, p. 75, 2019.
- [14] S. Trick, D. Koert, J. Peters, and C. Rothkopf, "Multimodal Uncertainty Reduction for Intention Recognition in Human-Robot Interaction," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7009–7016, Macau, China, July 2019.
- [15] L. Shi, C. Copot, and S. Vanlanduit, "Gazeemd, "Detecting visual intention in gaze-based human-robot interaction," *Robotics*, vol. 10, no. 2, p. 68, 2021.
- [16] A. Buerkle, W. Eaton, N. Lohse, T. Bamber, and P. Ferreira, "EEG based arm movement intention recognition towards enhanced safety in symbiotic Human-Robot Collaboration," *Robotics And Computer-Integrated Manufacturing*, vol. 70, Article ID 102137, 2021.
- [17] M. Liu, B. Peng, and M. Shang, "Lower Limb Movement Intention Recognition for Rehabilitation Robot Aided with Projected Recurrent Neural Network," *Complex & Intelligent Systems*, vol. 8, pp. 1–12, 2021.
- [18] K. A. Ingraham, C. D. Remy, and E. J. Rouse, "The role of user preference in the customized control of robotic exoskeletons," *Science robotics*, vol. 7, no. 64, Article ID eabj3487, 2022.
- [19] L. Roveda, B. Maggioni, E. Marescotti et al., "Pairwise preferences-based optimization of a path-based velocity planner in robotic sealing tasks," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6632–6639, 2021.
- [20] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [21] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, 2016.
- [22] J. Tjomsland, A. Shafti, and A. A. Faisal, "Human-robot Collaboration via Deep Reinforcement Learning of Real-World Interactions," 2019, <https://arxiv.org/abs/1912.01715>.
- [23] H. Bae, G. Kim, J. Kim, D. Qian, and S. Lee, "Multi-robot path planning method using reinforcement learning," *Applied Sciences*, vol. 9, no. 15, p. 3057, 2019.
- [24] F. Niroui, K. Zhang, Z. Kashino, and G. Nejat, "Deep reinforcement learning robot for search and rescue applications: exploration in unknown cluttered environments," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 610–617, 2019.
- [25] T. Yu, J. Huang, and Q. Chang, "Mastering the working sequence in human-robot collaborative assembly based on reinforcement learning," *IEEE Access*, vol. 8, pp. 163868–163877, 2020.
- [26] M. El-Shamouty, X. Wu, S. Yang, and A. Marcel, "Towards safe human-robot collaboration using deep reinforcement learning," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4899–4905, Paris, France, September 2020.
- [27] A. Ghadirzadeh, X. Chen, W. Yin, Z. Yi, M. Bjorkman, and D. Kragic, "Human-centered collaborative robots with deep reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 566–571, 2021.
- [28] L. Roveda, J. Maskani, P. Franceschi et al., "Model-based reinforcement learning variable impedance control for human-robot collaboration," *Journal of Intelligent and Robotic Systems*, vol. 100, no. 2, pp. 417–433, 2020.
- [29] R. Zhang, Q. Lv, J. Li, J. Bao, T. Liu, and S. Liu, "A reinforcement learning method for human-robot collaboration in assembly tasks," *Robotics and Computer-Integrated Manufacturing*, vol. 73, Article ID 102227, 2022.
- [30] Z. Liu, Q. Liu, L. Wang, W. Xu, and Z. Zhou, "Task-level decision-making for dynamic and stochastic human-robot collaboration based on dual agents deep reinforcement learning," *International Journal of Advanced Manufacturing Technology*, vol. 115, no. 11-12, pp. 3533–3552, 2021.
- [31] J. Ibarz, J. Tan, C. Finn, M. Kalakrishnan, P. Pastor, and S. Levine, "How to train your robot with deep reinforcement learning: lessons we have learned," *The International Journal of Robotics Research*, vol. 40, no. 4-5, pp. 698–721, 2021.
- [32] J. Guzzi, G. Abbate, A. Paolillo, and A. Giusti, "Interacting with a conveyor belt in virtual reality using pointing gestures," in *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 1194–1195, Sapporo Hokkaido, Japan, March 2022.
- [33] B. Gromov, G. Abbate, L. M. Gambardella, and G. Alessandro, "Proximity human-robot interaction using pointing gestures and a wrist-mounted IMU," in *Proceedings of the International Conference on Robotics and Automation (ICRA) IEEE*, pp. 8084–8809, Montreal, QC, Canada, August 2019.
- [34] S. Hopko, J. Wang, and R. Mehta, "Human factors considerations and metrics in shared space human-robot collaboration: a systematic review," *Frontiers in Robotics and AI*, vol. 9, Article ID 799522, 2022.

Research Article

Analysis on the Bus Arrival Time Prediction Model for Human-Centric Services Using Data Mining Techniques

N. Shanthi,¹ Sathishkumar V E,² K. Upendra Babu,³ P. Karthikeyan,⁴ Sukumar Rajendran,⁴ and Shaikh Muhammad Allayear⁵

¹Department of Computer Science and Engineering, Kongu Engineering College, Perundurai, Erode, India

²Department of Industrial Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea

³Department of Computer Science and Engineering, Bharat Institute of Higher Education and Research, Chennai, Tamil Nadu, India

⁴School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

⁵Department of Multimedia and Creative Technology, Daffodil International University, Daffodil Smart, Khagan, Ashulia, Dhaka 1207, Bangladesh

Correspondence should be addressed to Sathishkumar V E; sathishkumar@scnu.ac.kr and Shaikh Muhammad Allayear; drallayear.swe@diu.edu.bd

Received 23 March 2022; Revised 9 May 2022; Accepted 17 May 2022; Published 26 September 2022

Academic Editor: Zhongxu Hu

Copyright © 2022 N. Shanthi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The human-computer interaction has become inevitable in digital world. HCI helps humans to incorporate technology to resolve even their day-to-day problems. The main objective of the paper is to utilize HCI in Intelligent Transportation Systems. In India, the most common and convenient mode of transportation is the buses. Every state government provides the bus transportation facility to all routes at an affordable cost. The main difficulty faced by the passengers (humans) is lack of information about bus numbers available for the particular route and Estimated Time of Arrival (ETA) of the buses. There may be different reasons for the bus delay. These include heavy traffic, breakdowns, and bad weather conditions. The passengers waiting in the bus stops are neither aware of the delay nor the bus arrival time. These issues can be resolved by providing an HCI-based web/mobile application for the passengers to track their bus locations in real time. They can also check the Estimated Time of Arrival (ETA) of a particular bus, calculated using machine learning techniques by considering the impacts of environmental dynamics, and other factors like traffic density and weather conditions and track their bus locations in real time. This can be achieved by developing a real-time bus management system for the benefit of passengers, bus drivers, and bus managers. This system can effectively address the problems related to bus timing transparency and arrival time forecasting. The buses are equipped with real-time vehicle tracking module containing Raspberry Pi, GPS, and GSM. The traffic density in the current location of the bus and weather data are some of the factors used for the ETA prediction using the Support Vector Regression algorithm. The model showed RMSE of 27 seconds when tested. The model is performing well when compared with other models.

1. Introduction

Human-computer interaction is generally defined as “the science concerned with designing effective interaction between users and computers and the construction of interfaces that support this interaction.” The digital transformation has made computers inseparable from human life. Humans are dependent on digital gadgets like mobile phones, tablets, laptops, and so on to complete their day-to-day tasks. The increase in demand for human-computer interaction has

opened a wide scope of research in recent years. In terms of computer programming, many front end frameworks have been introduced to develop user-friendly mobile- or web-based applications. These applications help humans to simplify their regular tasks like paying bills, buying groceries or medicines, booking cabs, and so on. Another advantage is that even users with no technical background can also use these applications for their needs [1].

In India, the most used and convenient mode of transportation is the buses. Though this is the most used

transportation, its travelers do not know the information like what the arrival time and the exact routes and available bus in the route. Passengers are supposed to wait for an indefinite time not knowing about the bus they are waiting for and information regarding the bus. Because of lack of real-time information regarding bus location, route, and traffic information, it is hard for managing the bus services for drivers, travelers, bus managers, and policy makers. To address the problems faced by the bus user's bus managers and policy makers, an Intelligent Transportation framework is proposed [2]. Initially, a data collection module that collects information about routes, timestamp, real-time bus location, traffic information, and weather information is created. Once the data collection module is developed, by using Artificial Intelligence techniques, various useful patterns such as trip duration, arrival time, trip duration, transit pattern, bus scheduling, traffic management, and so on can be identified. All the possible information regarding the transportation is made transparent for the benefit of bus users, bus managers, bus drivers, and policy makers to keep informed about a trip. By developing such a transparent system, it is possible to reduce various problems, such as traffic congestion, waiting time, transit problem, scheduling problems, and so on. The main objective of this project is to develop a real-time bus management system using Artificial Intelligence for the benefit of bus users, bus drivers and bus managers which can effectively address the problems related to bus information system transparency [3].

The ITS application specifically aimed to develop for the Indian transportation system, as an easily accessible tool with various options for passengers, bus managers, and policy makers. Despite the availability of ITS applications, there are very few trial versions of transportation applications for traffic monitoring available and no specific tool combining all the tasks of ITS. An initiation is needed to develop a project combining all aspects of ITS in India [4]. If the idea of ITS is implemented, major problem of traffic congestion can be reduced in a considerable manner. Regardless in the developed countries, India needs ITS infrastructure development to tackle the upcoming technological advancements in ITS. India is land of various languages with several cultural practices. Proper information regarding the public transportation in all the states should be developed so that people traveling to a different city from their native city can use that information for their mobility. Because of the lack of public transportation information, travelers are facing lots of difficulties while traveling from one place to another. Peoples can be easily fooled by the false information conveyed to them by unknown authorities. So, an authentic public transportation information delivery system should be developed. This information will be more useful for passengers traveling in any part of the nation irrespective of their language barrier. As a first step, a data collection framework that collects all possible information from public vehicles, road and traffic scenario is required to develop an ITS system, which can provide transparent data about vehicle patterns to the bus managers, passengers, and policy makers so that it could be useful for planning, making

smart decisions, efficient management, and providing sophisticated public vehicle usage.

"Estimated Time of Arrival" (ETA) refers to the amount of time taken by a vehicle to reach its destination. It is a transportation concept that refers to the length of time it takes for any vehicle like bus, ship, helicopter, or emergency service to arrive at its destination [5]. ETA is commonly used to inform travelers about remaining time available before a certain mode of transportation reaches a particular destination. This paper proposes a real-time bus management system that use HCI-based web and mobile application to remind passengers about the estimated bus arrival time at their destination, considering various factors like traffic jams and weather conditions. In addition to ETA, this system also provides a list of all bus stops for a given bus.

Our contribution for this manuscript is listed below.

- (i) We established Intelligent Transportation System to collect, analyze, and identify patterns in the Indian transportation for the benefit of passengers, bus managers, and policy makers.
- (ii) We developed Information and Communication technology-enabled bus management system with advanced information processing system.
- (iii) We have created modules to track and display the movement of vehicles in real time under the influence of various factors, such as traffic, weather, and time parameters.
- (iv) We have recorded the daily trips and the routes including speed and traffic information with movements of vehicles in a series of topographical zones—geofencing.
- (v) We have developed module to deliver information regarding routes, arrival time, trip duration, traffic information, and transit patterns to the passengers.

The rest of the paper is organized as follows. The research papers related to this field are discussed in Section 2. The hardware and the software requirements are discussed in Section 3. The algorithm implemented in the paper is explained in Section 4. The implemented model is compared with other models and necessary illustrations are made in Section 5. The results and the scope for further research are discussed in Section 6.

2. Related Work

The Framework Program for Research, Technological Development and Innovation (DESMI 2008) of the Cyprus Research Promotion Foundation focuses more on Event-Based Bus Monitoring System (EBM) [6]. Specifically, the focus is on reducing contact signals in order to produce an acceptable result. Bus arrival times are meticulously tracked. EBM's findings indicate that it hires just 3.5 percent of the overall volume of signaling communications, bringing it down to a manageable level to a major degree.

M-ESB is a multisensor data collection and sharing interface for a handheld sensor grid and router that feeds

into a remote data cloud. It also introduces a new network business model in which the public bus company acts as a Virtual Mobile Service provider. The use of WSN technology as a method for managing traffic signals between Johannesburg and Pretoria is defined in Vehicle Traffic Monitoring Using Wireless Sensor Network in South Africa. Furthermore, RFID scanners are used in the system to detect congested areas and warn the traffic officer at the Traffic Monitoring System. The authors presented an integrated system [7] that monitors the current position of the bus/vehicle and indicates the actual position to the regular user, as well as alerting the regular user of any catastrophic event that may cause a natural slowdown during traveling [8]. This information about the bus's approach is stored on a back-end server, and commuters are informed of this information through a mobile application, allowing them to choose an alternative direction [9]. This device infers that the main emphasis is on the position of the buses and the potential delays due to any disaster [10].

Ingle [11] proposed a system aiming to develop a low-cost solution for helping the passengers obtain the information related to their buses and journeys by considering live bus location tracking, showing seating capacity, showing maximum number of standing passengers allowed in the bus, displaying number of passengers currently traveling in the bus, and calculating Estimated Time of Arrival as essential features of the system. The information about the bus fares from one place to another was also provided in the application. The Silver Cloud Real-Time GPS tracker was used in the system for tracking live location of the bus. The data were sent from the GPS tracker to the database through HTTP using POST method. The user interface was provided as a web application developed using JavaScript, MySQL, and Google Map APIs [12].

The designed methodology decreases the time that different users would wait for a bus. The bus can be tracked at any time and from any place using a device. All current data is saved on the server and accessed by remote users via a web-based application. This method allows users to get details directly displayed on a Google Map in a more user-friendly manner [13].

A smartphone application is used to monitor nearby vehicles and to send updates about them [14]. People can schedule their journeys and travel choices based on the proximity of bus stops [15]. It was introduced in order to change people's commuting decisions by taking into account the "Best Transport Division." [16]. The android application contains necessary details of the vehicle [17].

Luo et al. [18] proposed a framework based on IoT for public transport system integrating the bus, subway, and shared taxi, with their scheduling problems for proving better transfer solutions; additionally, methods are proposed for predicting the transport flow based on periodic patterns mining utilizing the passenger flow analysis and road flow analysis. A decision support system and mathematical model based evolutionary computation algorithm were used for dynamic bus scheduling and controlling problems. This IoT-based system can assist the passengers to utilize the transportation systems effectively and can reduce the travel time.

Chavhan et al. proposed an Internet of Things-based Intelligent PTS (IoT-IPTS) in a metropolitan region [19]. IoT is utilized to interconnect transportation elements, like vehicles, routes (sensors), commuters (cell phones), and side of the road units in a metropolitan territory. The IoT gives consistent connectivity between various networking systems at whatever point the passengers or vehicles move starting from one area to the next area. Subsequently, IoT gives the reasonable and consistent public transportation administrations in the metropolitan territory. Moreover, context data of transportation elements, for example, condition of routes, traffic congestion, number of routes accessible, vehicles movement pattern, and mobility, are stored in the cloud. The stored data in cloud alongside the IoTs are utilized to locate the significant routes, alternative modes, arrival time, departure time, transit planning, and many more for giving public transportation administrations in a metropolitan zone.

Chavhan et al. proposed dynamic vehicle allotment framework for public transportation using Emergent Intelligence (EI) strategy in a metropolitan zone. Also, the EI method's ability for tackling public vehicle framework issues is illustrated. EI method keeps up recorded data, commuters' appearance rates, availability of resources, and deficit in resources; an EI strategy is used to gather, investigate, share, and ideally assign transport resources adequately [20].

Treethidaphat et al. utilize GPS information from a public transportation line for bus to build up a bus arrival time forecast at any distance along the considered route [21]. Deep learning is utilized to get high accuracy. The accuracy of the model is assessed by real-time BMTA-8 bus transport information in Bangkok, Thailand, and contrasted the outcome with ordinary least square regression model. The result shows that the proposed deep learning model is more precise than the ordinary least square regression model around 55% for mean absolute percentage error. This shows that deep learning could be used as an effective tool for predicting arrival time [22].

Peilan et al. proposed a prediction strategy using bus trajectories, considering bus route and road network [23]. Passengers' multiple trips were accounted for, including waiting time at multiple points. A deep learning-based model is developed for multiway travel time prediction. Various experimental approaches were proposed to validate the supremacy of considered multiway dataset considered. Using an offline bus location data, also some studies were done to predict the bus running time [24]. Based on the prediction, bus schedules were updated.

Even though some of Intelligent Transportation projects are implemented in some cities in India, all these projects are small-scale standalone pilot studies. Even though they are not of integrated nature, significant efforts that have been made for employing ITS in various cities are discussed. It is evident from this scenario that there are several avenues available for ITS application to flourish, in spite of its growing popularity among the transport authorities. Also, it demands a systematic approach. After the ITS application is implemented at road network level, its complete benefits can be seen. It cannot be seen at the small scale or corridor level.

On viewing the present transportation context in India, emergency management, congestion management, advanced traffic management systems, advanced traveler information systems, commercial vehicle operations, advanced vehicle control systems, and so on are the aspects that require focus, other than the existing ITS applications.

In all these researches, the systems developed did not use traffic density and weather as factors for influencing the calculation of arrival time. These two factors are the major parameters used as input attributes for the machine learning model to train and predict the Estimated Time of Arrival of the bus in real time.

3. Methodology

3.1. System Architecture. The major components of the hardware kit comprise of the following:

- (i) Location provider module
- (ii) Centralized real-time database
- (iii) HCI interface

3.2. Location Provider Module. The location provider module consists of a GPS module, a GSM/GPRS module, and an IoT controller like Arduino or Raspberry Pi. Here, Raspberry Pi is used for quick development and easier debugging. Figure 1 presents the overall system architecture and Figure 2 shows the prototype built using the architecture in Figure 1.

3.3. GPS-Raspberry Pi Configuration. Once the GPS module is connected to a power source, the module tries to get a position fix with the help of the antenna that searches for any nearby satellites to get the current location of the module. The red led starts to blink once the module gets its location fix. After getting the location fix, the module starts transmitting location and other data in the form of an internationally standardized string that is standardized by NMEA (National Marine Electronics Association). This NMEA formatted GPS data contains number of NMEA messages that represents each type of data transmitted by the satellites. The data is transmitted to the controller module where the NMEA messages get parsed and processed to actual location data.

3.4. GSM/GPRS-Raspberry Pi Configuration. Before configuring the GSM/GPRS sim 900 A module with Raspberry Pi, a sim card with sufficient data should be inserted into the sim Module. The sim Module has all the functionalities similar to the features of the mobile phone, like calling, messaging, using Internet, and so on. These functionalities are carried out by AT commands that are passed to the sim Module from the Raspberry Pi controller. These commands instruct the sim Module to turn on data when connection is made with the controller. The ground of the Sim 900 A module is connected to one of the ground pins of Raspberry Pi. The power source of this module is a 12 V DC power

supply that can be supplied either with a DC adapter charger to a wall plug or a DC 12 V battery.

3.5. Real-Time Database. As the name suggests, a real-time database works as a state machine, triggering certain events when the state of the database, more precisely when a data in the database gets added, updated, or deleted. When the data gets updated, the database synchronizes the data with the clients' local data. In this way, the database sends the updated data to the client instead of client sending a new request every time to the database to acquire the updated data. Google's Firebase Real-Time database is the perfect example of this functionality. The database works under the principle of data synchronization with the clients connected to the database and the data is stored in a nonrelational JSON tree. Each data represents a node in the JSON tree and each node can act as a root node and can have its own tree. This feature of the real-time database is used in a manner where the database sends the response to the connected client whenever the data changes in the database.

3.6. HCI Interface. The client interface is a progressive web application that is built using AngularJS. The client shows the current bus stops and the other stops it crossed including the estimated time of arrival of the bus at each stop. The interface consists of three components: a map to show the current bus route and location markers, a side-menu to choose different bus routes, and a sidebar to see the ETA for different stops of the current bus.

The client is connected with the real-time database where the data is synchronized with the client on every update in the database. The map interface is taken from the Google Maps API that provides the location markers and the map of the route of the current bus that is selected. The bus stops are manually created in the database for each bus called as waypoints, where each waypoint denotes a bus stop.

3.7. Deployment in Bus. The hardware module shown in Figure 2 was deployed in a Kongu Engineering college bus, route number 82. The route 82 has the longest route among all buses running in the college; the source point is the college in Perundurai, and the destination is Sankagiri post office. The route map is shown in Figure 3.

3.8. Control Flow. The data and control stream starts with the GPS module and ends with the client interface. The control flow diagram given in Figure 4 helps in determining how the data flows throughout the modules. Once the GPS module gets the location fix, it starts sending the NMEA messages to the Raspberry Pi through the TX pin connected to one of the GPIO pins of Raspberry Pi. The messages are read using PyGPIO python library that reads serial data from the GPIO pins. The controller then parses the NMEA messages to fetch the respective latitude and longitude coordinates of the current location given by the GPS module. The parsed location data is sent as a request to a third-party traffic and weather API that fetches the traffic density and

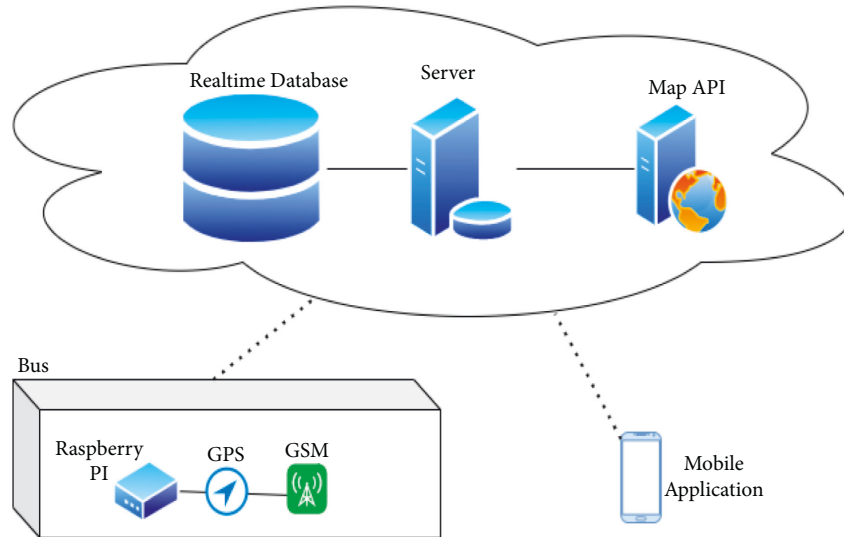


FIGURE 1: System architecture.

weather data of that location. The fetched traffic data and weather data are then processed by the controller and sent to the real-time database as nonpreprocessed raw data. For the third-party API, HERE Maps API is used to get both weather and traffic data for the location given by the Raspberry Pi controller. Whenever the vehicle crosses one of the waypoints, the controller computes the average traffic density between all crossed waypoints and the distance between the last crossed waypoint and the current way point. The duration between the two waypoints is also taken into account. These data along with the current waypoint location and the crossed waypoint location and names of those waypoints and the time at which the vehicle crossed the last waypoint with the time at which the vehicle arrived to the current waypoint makes up for the ML data that is sent to the real-time database. The accuracy of the location provider module is adjusted for getting the best results in terms of the radius under which the GPS gets the location coordinates and the controller parses the data and the GSM/GPRS module sends the data to the real-time database. The accuracy ranges from 10 meters to 250 meters maximum.

All these data are parsed into JSON objects and serialized to normal strings and sent as each request to the firebase real-time database and to the third-party HERE Maps API. The response contains the weather data of the current location in JSON format. Changing these weather data types into numeric factor can help in training the model.

This weather data is used as one of the factors for training the ML model. Another important factor is the traffic density the HERE Maps API provides. This traffic density varies in the range of 0 to 10 where “0” is the chosen location completely free of traffic and “10” is the chosen location completely full of traffic jams and heavy traffic. This traffic density has a great influence over the ML model that is being trained for the prediction of ETA of the vehicle. After collecting the required data from the HERE Maps API, the JSON responses are appended to the processed ML data and stored in the firebase real-time database. This data acts as

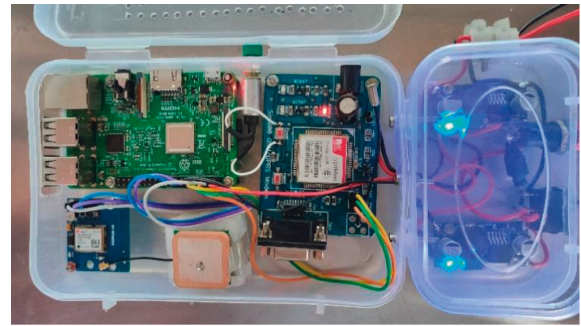


FIGURE 2: Real-time vehicle tracking module.

input to the ML model for training and testing. The predicted ETA is then stored in the real-time database for the client interface.

The HCI interface fetches the ETA of the current selected vehicle from the list of routes and displays it to the user. Here the change in ETA triggers a change in state of the data in the real-time database which in turn triggers the responses to be sent to all the clients connected with the real-time database. The interface gets updated data every time the real-time database synchronizes the data with the client. As each route is selected from the client interface, the client sends a request to fetch the data from the real-time database. As the vehicle moves, the current location gets updated in the real-time database and triggers data synchronization with the client. The client displays the current location of the vehicle in map with predicted ETA for the next arriving waypoint using the help of Support Vector Regression model.

3.9. Support Vector Machine. Support vector machine is a simple classification algorithm that is used to find a hyperplane in an N -dimensional space, where N is the number of features that distinctly classifies the data points. The input data can be projected to higher dimensions if the data cannot be separated at lower-dimensional space. This can be

achieved using Kernel functions. Being able to learn from very small datasets, avoiding local minimums and generalization capability are some advantages of support vector machine [25].

Hyperplanes are the decision boundaries created by the algorithm to classify the data points into different class labels. Margin is the distance between hyperplane and the data points of each class. Kernel function is the function that takes data points as input and transforms them into required form. The different kernel functions that are available are linear, polynomial, radial basis, sigmoid, and so on. It returns the inner product of two data points in a feature space [26].

The support vector machine algorithm finds the equation for the optimal hyperplane from the training data and uses it for later predictions. The confidence value for the classifications will be directly proportional to the distance between the current data point and the hyperplane, which is the decision boundary [27]. The main purpose behind finding this optimal hyperplane that is far from all data points is to maximize the confidence value for future predictions. Consider there are input vectors x_i , where i ranges from 1 to n , reflecting the number of features that affect the result of the algorithm, weight vectors w_i that is the linear combination to classify the class labels or predict the value of y in case of regression and b for intercept. The value for y used by support vector machine is $y \in \{-1, 1\}$ in case of binary labels. For each training example (x_i, y_i) , a corresponding functional margin Y_i of (w, b) is used to check for the confidence value of each prediction. Here, w^T is the weight vector and b is bias term. It can be mathematically seen as given in the following equation:

$$Y_i = y_i(w^T x + b). \quad (1)$$

As already mentioned above, the confidence value will be higher if the margin value is higher. If y_i is 1, then the functional margin value should be large, which in turn depends on $(w^T x + b)$ being larger and positive based on the above equation. In contrast, if y_i is -1, then $(w^T x + b)$ should have larger magnitude but be negative [28]. Therefore, for the most accurate predictions, the functional margin value must be larger than 0 and as close as 1. If it is greater than 0, the prediction might be correct.

$$Y_i > 0 \Rightarrow y_i(w^T x + b) > 0. \quad (2)$$

To find the decision boundary that maintains maximum distance with each data point of each class, the magnitude should be taken into consideration since it is not possible to calculate the distance without the magnitude. The functional margin is normalized with the Euclidean distance d , which is the distance between the data point and the decision boundary. Hence, (1) can be modified as below:

$$Y_i = y_i \frac{(w^T x + b)}{\|w\|}. \quad (3)$$

By dividing the functional margin by the magnitude, a constraint is posed on the size of w that maintains the same value for high values with same proportions, thus

normalizing the functional margin. So, the geometric margin that is the Euclidean distance and the distance between the decision boundary and positive boundary can be given as the following equation:

$$\frac{(w^T x + b)}{\|w\|} = \frac{1}{\|w\|} \text{ and } \frac{2}{\|w\|}. \quad (4)$$

To maximize $2/\|w\|$ is the same as to minimize $1/2 * \|w\| * \|w\|$, which is a quadratic program that can be solved easier. Hence, the following minimization condition can be given:

$$\text{Minimize } \frac{1}{2} * \|w\| * \|w\| \text{ such that } y_i(w^T x + b) - 1 \geq 0. \quad (5)$$

This problem can be solved using Lagrangian multiplier method since it is constrained quadratic optimization problem by posing a multiplier on the constraint. This multiplier is called as Lagrangian multiplier and this makes the equations at any data point; when the support vector is not present, the value for route function a_i and a_j becomes zero. The equation is given as

$$\text{Min}(w, b, \alpha) = \frac{1}{2} \|w\|^2 * \frac{1}{2} * \|w\|^2 - \sum_{i=1}^m a_i y_i (w^T x + b) \dots 1. \quad (6)$$

Deriving the Lagrangian multiplier from equations (4)–(6) will produce the below equations:

$$w = \sum_{i=1}^m a_i y_i, \quad (7)$$

$$\sum_{i=1}^m a_i y_i = 0.$$

Maximizing over a can produce new Lagrangian equation by getting rid of dependence on w and b with replacing w in the above equations. Thus, the below shown dual optimization problem can be produced.

$$\text{Max}(a) = \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m y_i y_j a_i a_j \langle x_i, x_j \rangle \text{ such that } a_i \geq 0, \quad (8)$$

$$\sum_{i=1}^m y_i y_i = 0.$$

The above equation shows that the dot product of x_i and x_j influences the maximization of the “ a ” value. When the inner product of x_i and x is large and from different classes, it forms the margin with maximum width whereas the inner product of the same class does not yield any significance [29]. The value of w and b can be in turn obtained by obtaining the a -value that maximizes $L(a)$. Thus, the final equation can be given as

$$w^T + b = \left(\sum_{i=1}^m a_i y_i x_i \right) T x + b = \sum_{i=1}^m a_i y_i \langle x_i, x \rangle + b. \quad (9)$$

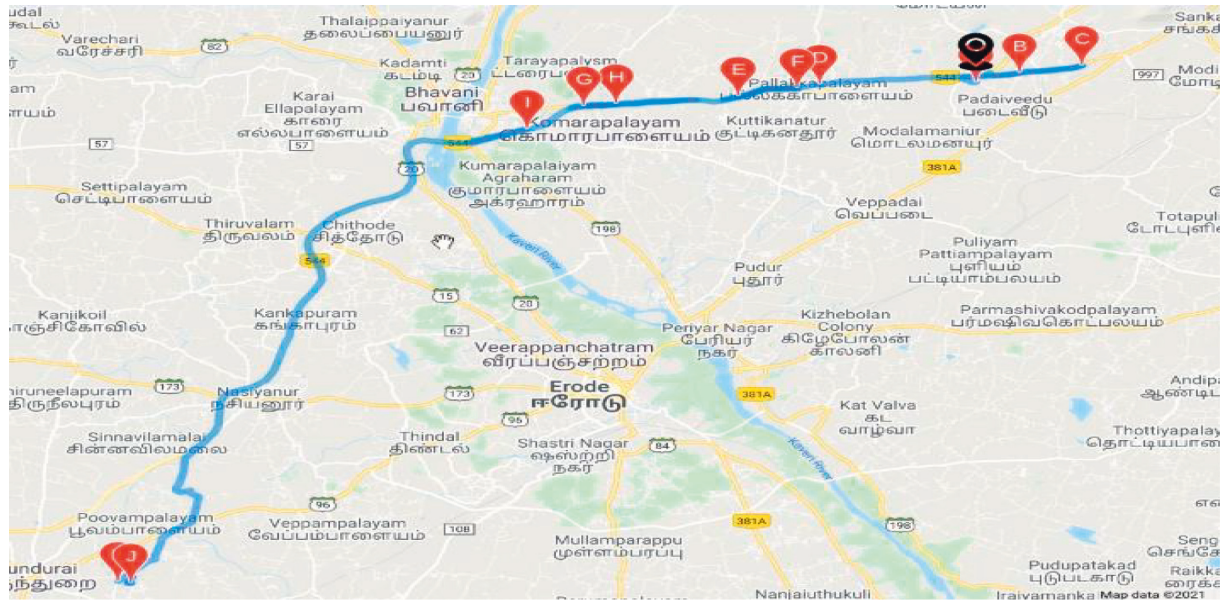


FIGURE 3: Route 82 roadmap.

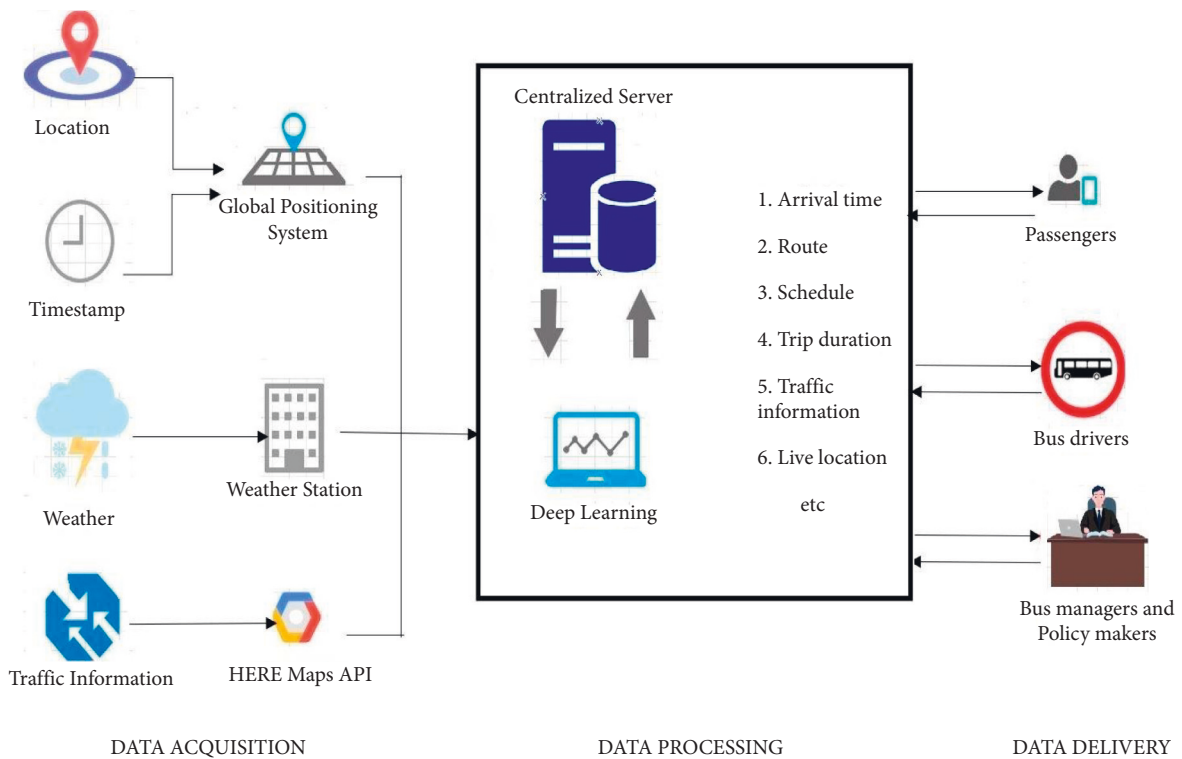


FIGURE 4: Data flow diagram.

Since the “a” value will be zero for all points except the support vectors, the summation of the classifier equation for all non-support vectors will result in zero. So, the prediction is done by considering only the inner product of support vector and the newly provided x which may result in values greater than or equal to zero or less than zero. Greater than or equal to zero means the prediction class is positive and less than zero means the prediction class is negative [30].

Support vector regression works similar to support vector machine and uses the same principles but for regression problems while support vector machines are widely used for classification problems. Support vector regression is not a probabilistic approach and does not assume any randomness. The main objective of the support vector regression algorithm is to find a function that approximates mapping from the input data points to a real number on the

basis of the provided training sample. The function defines the hyperplane that can be of any shape and mostly high dimensional based on the number of features affecting the result to produce minimum error. This algorithm also uses concepts used in support vector machine, like the hyperplane, decision boundary, and margin. The same constraints that are used in the support vector machine are used for support vector regression. The constant distance d from the hyperplane to the nearest data points defines the equations for decision boundaries. The equations are given as $wx + b = +d$ and $wx + b = -d$ for positive and negative decision boundaries, respectively. Thus, the equation of any hyperplane that satisfies the support vector regression should satisfy the below equation:

$$-d < y - wx + b < +d. \quad (10)$$

The algorithm basically considers the data points that are within a decision boundary and produces the best-fit line or plane, which is the hyperplane that has maximum number of points. The margin tolerance C and the decision boundary distance ϵ can be manipulated by the supervisor for better results with minimum error and producing a better fitting model. From this, support vector regression provides the user the flexibility to define how much error is acceptable in the model and find an appropriate hyperplane to fit the data.

4. Developing the Model

The data obtained from the real-time vehicle tracking device is stored on a real-time database. These data are pre-processed and converted into the dataset with the attributes like source location, destination location, distance, average jam factor, average weather, and duration of travel. All these input features are used in training the support vector regression model. The training set contains 75 percentages of the data, and the test set contains the remaining 25 percentage. The transportation data is collected for 30 days. A total of 3360 entries are recorded. Among these, 2520 are used for training the model and 840 are used for testing. Here, the model is trained to provide the duration of travel as a result from the other input features. The model is then integrated with a web server to use the predictions for real purposes.

4.1. Data Preprocessing. The data obtained from the real-time vehicle tracking module is the raw data. The data contains only information about the current location, like latitude, longitude, timestamp, jam factor, and weather. The data is then analyzed, and the dataset is generated. Python scripts are written for detecting the data produced at the waypoints of the bus. These data are separated, and another script is written to transform this data into a dataset that contains source, destination, distance, average jam factor. The average jam factor is calculated by taking average of all the jam factors of the data instance generated from source to destination. The average weather is calculated similar to the average jam factor. The duration of travel is calculated by

taking difference between the timestamp of the data instance from source and the destination [31].

The dataset is ready now from the theoretical point of view, but to use this dataset with the python functions, the data cannot be strings and all the instances in the dataset should be converted to float values. For this purpose, scikit-learn provides Label Encoder class for converting all the string values into float values. And since all the values present in the dataset may not be at the same range of numbers, they need to be transformed so that all the input features fall in the same range of numbers. For example, the duration is in seconds, so the values can be in thousands range if the bus route is long enough, but the jam factor values range between 0 and 10. It may lead to difficulty in plotting the data points.

This problem is resolved by a technique called feature scaling, where all the input features are transformed to the same range. There are various methods for feature scaling. The min-max scaling is used here. The mathematical formula for min-max scaling is given as

$$x_i^n = \frac{(x_i - \min(x_i))}{(\max(x_i) - \min(x_i))} \quad (11)$$

Here, x_i^n is the scaled value, x_i is the i th value, and min and max are the minimum and maximum values for the given range. The raw dataset and the processed dataset are shown in Tables 1 and 2. A 3D visualization of the dataset is shown in Figure 5.

4.2. Building and Training the Model. The model is developed and trained in the python programming environment using various libraries, like Pandas, Scikit-Learn, NumPy, and Matplotlib. The data from the real-time database is cleansed and converted to a CSV file. The Scikit-Learn module provides the functions for implementing the support vector regression algorithm.

The preprocessed data is separated into two datasets in the ratio of 3:1 using `test_train_split` function provided by the Scikit-learn library. The larger part is fed to the support vector regression model whose implementation method is provided by Scikit-learn library. The model training can be manipulated using four different parameters. The process of changing these parameters is called hyperparameter optimization [32]. The first parameter is C , which defines the weight of how much samples inside the margin contribute to the overall error. This allows the user to optimize both the fit of the line to data and penalize the samples inside the margin. It in turn allows the user to adjust how hard or soft the margin classification should be. With high values of C , the samples inside the margin are penalized more. The second one is epsilon. Epsilon defines the value of margin where the errors are tolerated and not penalized. The larger the value of epsilon, the larger the number of errors admitted. The third parameter is kernel. Kernel is the function that takes input data and transforms it into the required form of processing data. Kernel function transforms the training dataset so that a nonlinear decision surface is able to be transformed into linear equation in a higher-dimensional

space. The return value of this function is the inner product of two data points in a standard feature dimension.

There are different types of kernel functions available like linear kernel that is used when the data points are linearly separable, Gaussian kernel which is used when there is no prior knowledge about the data points, radial basis function which is the most commonly used, sigmoid kernel, polynomial kernel, and so on. The fourth parameter is the gamma. Gamma is the hyperparameter that decides how much curvature the decision boundary can have. The magnitude of gamma is directly proportional to the curvature of the decision parameter.

The values for these parameters used for training the model are chosen by running a python script that trains the model with a range of values for C and epsilon. The technique implemented in the script is the grid search technique. The gamma is auto, which considers $1/n$, where n is the number of features and kernel function is radial basis function. These two parameters were constant for all pairs of c and epsilon. The model trained with different values of c is tested for error. The model with the lowest error is taken for training the final model. Then, epsilon is changed within a range, where c is the resultant of previous step and the model with lowest error is found. This value of epsilon is used for final model. The final model is trained using the C and epsilon values that resulted in the previous process and the other parameters remain as default.

5. Testing the Model

The trained model is tested for accuracy with one-quarter of the dataset. Unlike classification problems, regressions do not produce absolute binary values but rather they provide a numeric value in a range as a result. There are various metrics to measure the errors produced by an algorithm.

5.1. Mean Squared Error (MSE). This method measures the average of squares of the error, which means average of the differences between the actual value and the predicted value. The result will always be nonnegative. The results closer to zero are better.

$$MSE = \frac{\sum_{i=1}^n (AT_i - PT_i)^2}{n} \quad (12)$$

Here, AT_i is the actual time, PT_i is the predicted time, and n is the number of predictions.

5.2. Mean Absolute Error (MAE). This is the simplest error measurement method where the error is calculated as the average of the absolute differences between the actual values and predicted values. It is mathematically given as

$$MAE = \frac{\sum_{i=1}^n (AT_i - PT_i)}{n} \quad (13)$$

Here, AT_i is the actual time, PT_i is the predicted time, and n is number of predictions.

5.3. Root Mean Square Error (RMSE). This is a quadratic-based rule to measure the absolute average magnitude of the error. It is calculated by summing all the differences between actual values and the predicted values, squaring the difference value and dividing the sum with the number of predictions and finally taking a square root of the value. Since the values are squared and rooted, the result will always be positive. The mathematical formula is given as

$$RMSE = \sqrt{\frac{\sum_i^n (AT_i - PT_i)^2}{n}}, \quad (14)$$

where AT_i is the actual time, PT_i is the predicted time, and n is number of predictions.

5.4. Relative Absolute Error (RAE). This method is similar to the mean absolute error, but instead of using actual value, a simple predictor is used to provide with values in place of actual values.

5.5. Relative Squared Error (RSE). This method compares the model with a simple predictor. Total squared error of the tested model is normalized and divided by total squared error of the simple predictor.

This trained prediction model is tested for accuracy using Mean Squared Error and Root Mean Squared method. And grid sliding technique is employed to find the least error possible hyperparameters.

6. Comparison with Other Models

The same dataset was used for training and testing different models so that we can compare the performance of SVR with others. The algorithms chosen for other models were Random Forest Regressor, Decision Tree Regressor, K -Neighbors Regressor, Gradient Boosting Regressor, XGB Regressor, and AdaBoost Regressor. The results produced by the models are shown in Table 3. Graphs of actual duration against predicted duration were plotted for each of the models. The linearity in the graph shows the accuracy of prediction. So, if the graph is as linear as the line $y = x$, then the model works with 100 percent accuracy. The actual by predicted plot is a scatter plot. The predicted response (\hat{Y}) is used for the abscissa. The observed response (Y) is used for the ordinate. The plot can also be used to visually evaluate the possibility of "lack of fit." An unbiased prediction should produce predicted values that agree with the observed values on average. If the model is biased, then the data points will deviate from the line. For example, if the response to changing the factors is nonlinear but the model includes only terms for linear effects, then the model will be biased. Figures 6–12 show the predicted versus actual plot for SVR, RFR, DTR, KNNR, GBR, XGBR, and AdaBR, respectively.

Figures 13–15 represent the comparison of R -squared, mean squared error, and root mean squared error for regression algorithms considered. All the data shown in the table are visually represented in the bar graphs. The y -axis in

TABLE 1: Raw dataset.

Latitude	Longitude	Date	Time	Jam factor	Weather
11.45428433	77.8137363	11-03-2021	07:21:06	1.02058	Passing clouds
11.45513267	77.81313	11-03-2021	07:33:26	1.80532	Passing clouds
11.27909933	77.59135	11-03-2021	08:19:52	0.02995	Fog
11.27546733	77.58735	11-03-2021	16:58:09	2.50171	Partly sunny
11.45908733	77.839009	11-03-2021	17:59:38	3.92654	Scattered clouds

TABLE 2: Processed dataset.

Source	Destination	Average jam factor	Average weather	Duration (seconds)
Goundanoor	ICL PO	2.33693	Scattered clouds	309
Goundanoor	KPR school	3.07268	Scattered clouds	245
Katheri road	Kottaimedu bus stop	1.389322	Passing clouds	186
Katheri road	Kottaimedu bus stop	4.66105333	Scattered clouds	254
KPR school	Muniyappan Kovil	0	Scattered clouds	345
KPR school	Muniyappan Kovil	0.017798333	Scattered clouds	350

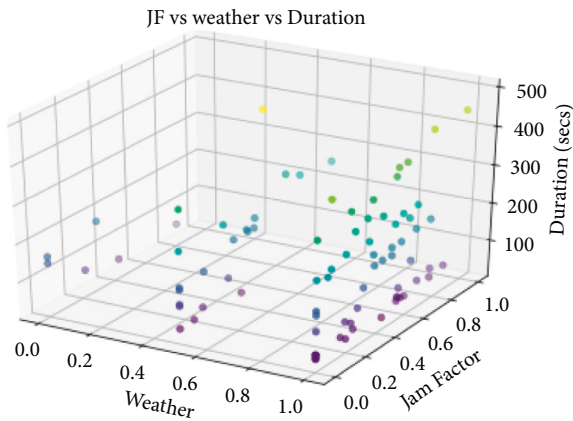


FIGURE 5: 3D Visualization of dataset.

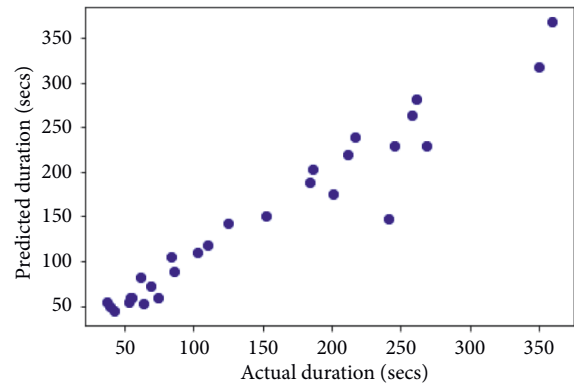


FIGURE 6: Predicted versus actual plot for SVR.

TABLE 3: Comparison of different models.

Algorithm	R-Squared	MSE	RMSE
SVR	0.914788192823	433.41898834	20.8187172
RFR	0.721238437404	2221.3996912	47.13172701
DTR	0.597991021123	3203.5357142	56.59978546
KNNR	0.842112238521	1258.17857142	35.47081295
GBR	0.38344528449	4913.21128311	70.09430278

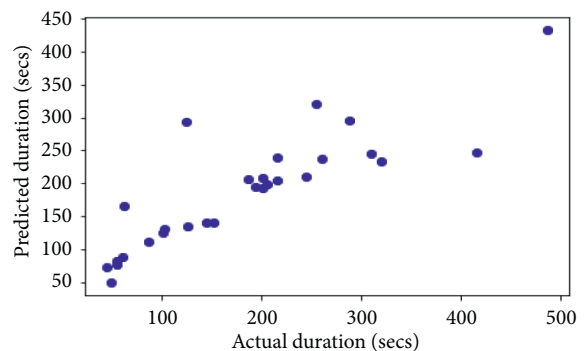


FIGURE 7: Predicted versus actual plot for RFR.

those bar graphs represents error value by the respective error measuring metric and the x -axis represents the different models measured for error. From all the scattered point graphs of predicted values against actual values, the model with support vector regression algorithm produces results quite resembling the linear graph whereas in other models the points were a bit more scattered. From Table 3, it is clear that support vector performs the best among all the taken models with an accuracy of 92 percentages. K -Neighbors Regression was kind performing nearly as good as Support Vector Regression with only 15 seconds of more RMSE than SVR. But when it comes to real-time usage, the

model might have to predict the ETA continuously while the buses are being operated. So, choosing the KNNR algorithm will result in a lot of computation and prediction will be very slow since it is a lazy algorithm and storing all training data would require more memory. Random Forest regression was the better performing model next to KNNR with an RMSE of 47 seconds. This RMSE is more than twice of that of SVR. The usage of Random Forest regression for this case would

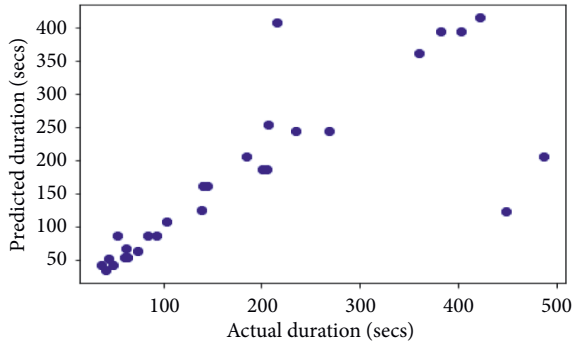


FIGURE 8: Predicted versus actual plot for DTR.

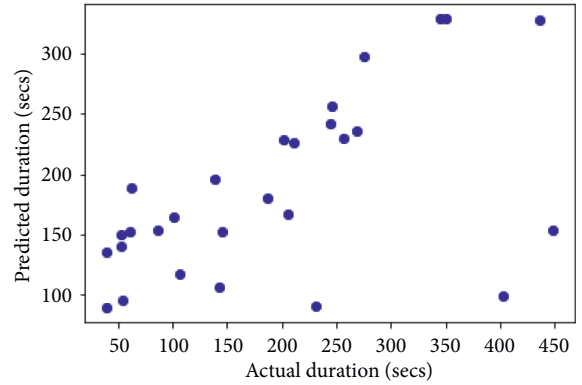


FIGURE 12: Predicted versus actual plot for AdaBR.

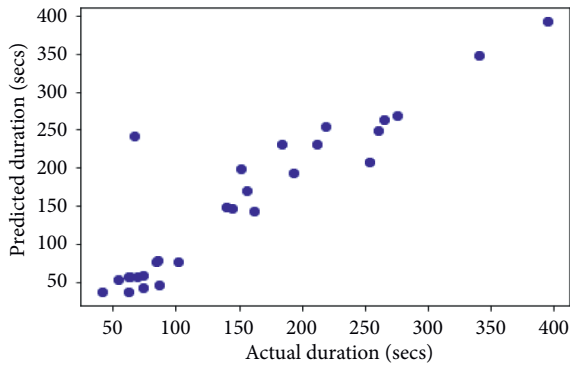


FIGURE 9: Predicted versus actual plot for KNNR.

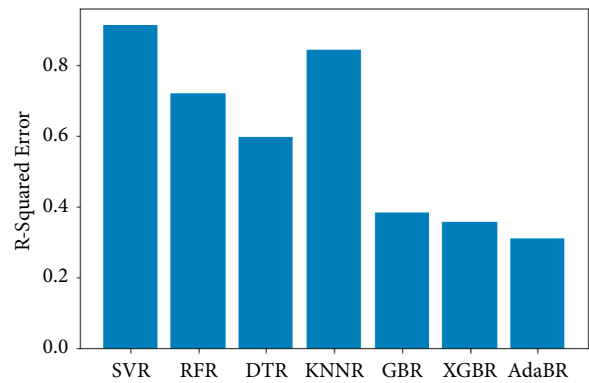


FIGURE 13: Comparison of R-squared error.

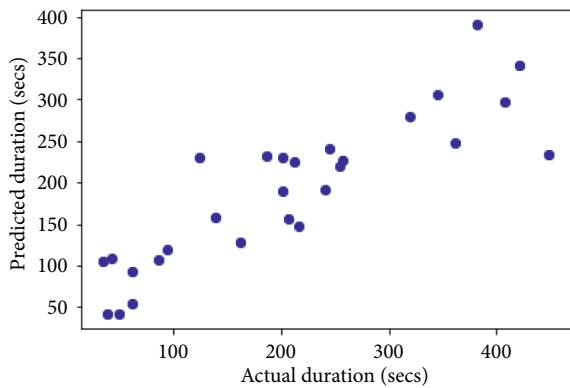


FIGURE 10: Predicted versus actual plot for GBR.

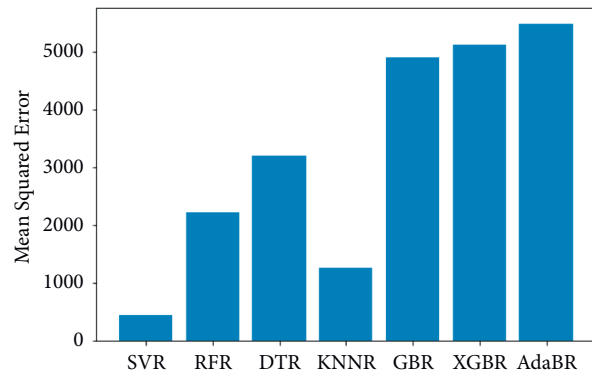


FIGURE 14: Comparison of mean squared errors.

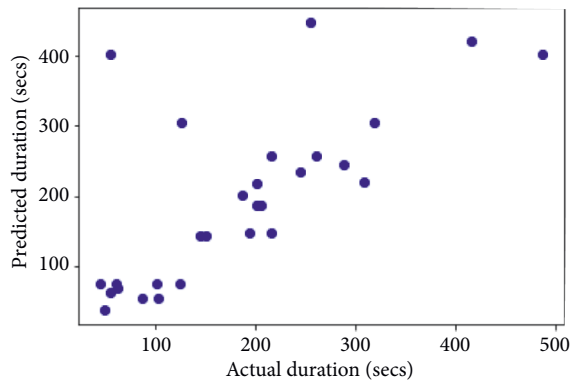


FIGURE 11: Predicted versus actual plot for XGBR.

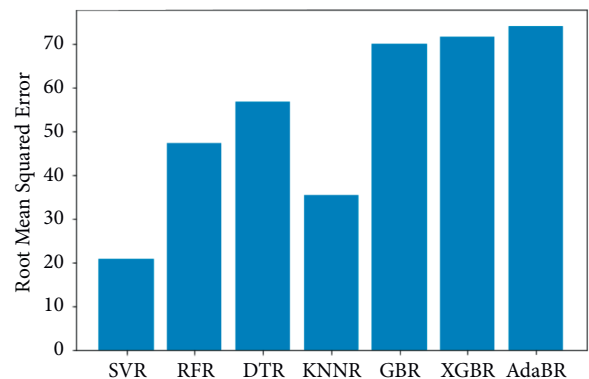


FIGURE 15: Comparison of root mean squared errors.

have been better if the data set consisted all possible values for every feature. Since the dataset is being generated in real time, a data instance with certain weather and a certain jam factor may not even be obtained before the situation occurs. That means a jam factor of value 10 will occur when the road is fully jammed and filled with vehicles where nothing can move. This situation never occurred while gathering the data. So, when using RFR, the model may not be able to extrapolate that may lead to very poor predictions. The other models did not perform well and the RMSE were very much higher compared to SVR.

7. Result Analysis and Conclusion

After testing the model with one-fourth of the total dataset, the model was producing results with Root Mean Square Error of 20 seconds. The accuracy of the predictions is better when the RMSE value is as low as zero. The error shown by the model is less than 0.5 minutes and that is very low compared to the duration of the travel time of the bus. The travel time of the bus deployed with the remote bus tracking module is nearly one and half hours, which is 90 minutes. So, comparing the error with the duration, the error is very much small. It is less than 0.1 percent of the duration. But when compared to the duration between waypoints, it will be 6–8 percent since the average duration between waypoints is 6–7 minutes. This result was achieved by tuning the hyperparameters of the Support Vector Regression algorithm using the grid search technique.

This model considers the jam factor and weather conditions, which are not considered in other models. From similar researches, it can be observed that other models have 40–50 seconds of error for each station. Comparatively, this model performed better with only 20 seconds RMSE. The Support Vector Regression algorithm implemented here predicted Estimated Time of Arrival (ETA) accurately and the real-time bus tracking module facilitated gathering the information effectively.

There is still some room for improvements. The first one would be using high configuration components for the real-time bus tracking module like better GPS module and GSM that supports 4G data communication. The second improvement would be reducing the load of the Raspberry Pi by making all the API calls from separate server so that Raspberry Pi only focuses on transmitting the location data. Next improvement would be using bigger dataset for training the model to achieve better performance from the model. This improvement can be made easily but requires quite a lot of time to generate all the real-time data.

Both mobile- and web-based application will be developed to predict the trip duration of the buses. These applications will be hosted in the website so that bus travelers, bus drivers, and bus owners can use the app and accurately predict the location of the buses. It is also proposed to give an alarm if the bus is struck with an accident or traffic jam. Any user with less proficiency can also use the system as it is a simple easily accessible system. The dataset collected will be published online for future research purpose.

The major challenge in real-time implementation will be the networking between all the modules, the database, and users. The next challenge will be performance tuning, when using the same model for predicting in all routes. This paper can be implemented as a real-time application with some changes like performance tuning in the model, training the model with more data, and providing users with additional features in the website and a dedicated mobile application. This can also be extended for other transportation systems. It can be very useful in industries for supply chain management.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

References

- [1] S. Rajendran, S. K. Mathivanan, M. S. Somanathan et al., "Detection and localisation of cars in indoor parking through UWB radar-based sensing system," *International Journal of Ultra Wideband Communications and Systems*, vol. 4, no. 3-4, pp. 182–190, 2021.
- [2] V. E Sathishkumar and Y. Cho, "A rule-based model for Seoul Bike sharing demand prediction using weather data," *European Journal of Remote Sensing*, vol. 53, no. 1, pp. 166–183, 2020.
- [3] M. Venkatesan, P. Mani, P. J. Kumar et al., "Remote monitoring of indoor/outdoor movement in epidemiological situations utilising UWB transceivers," *International Journal of Ultra Wideband Communications and Systems*, vol. 4, no. 3/4, p. 124, 2021.
- [4] V. E Sathishkumar, J. Park, and Y. Cho, "Seoul bike trip duration prediction using data mining techniques," *IET Intelligent Transport Systems*, vol. 14, no. 11, pp. 1465–1474, 2020.
- [5] K. B. Priya, M. S. Kumar, M. Geetha et al., "Queueing network model with jockeying to reduce the waiting time in the airport," *International Journal of System of Systems Engineering*, vol. 11, no. 3/4, p. 363, 2021.
- [6] W. Fan and Z. Gurmu, "Dynamic travel time prediction models for buses using only GPS data," *International Journal of Transportation Science and Technology*, vol. 4, no. 4, pp. 353–366, 2015.
- [7] S. Ve and Y. Cho, "Season wise bike sharing demand analysis using random forest algorithm," *Computational Intelligence*, vol. 1, 2020.
- [8] M. S. Kumar and J. Prabhu, "A hybrid model collaborative movie recommendation system using K-means clustering with ant colony optimisation," *International Journal of Internet Technology and Secured Transactions*, vol. 10, no. 3, p. 337, 2020.
- [9] V. E. Sathishkumar, P. Agrawal, J. Park, and Y. Cho, "Bike sharing demand prediction using multiheaded convolution neural networks," *Basic and Clinical Pharmacology and Toxicology*, vol. 126, pp. 264–265, 2020.

- [10] Z. Gurmu and W. Fan, "Artificial neural network travel time prediction model for buses using only GPS data," *Journal of Public Transportation*, vol. 17, no. 2, pp. 45–65, 2014.
- [11] D. Ingle, "Experimental estimates of low-cost bus tracking system using area-trace algorithm," in *Proceedings of the 2015 Fifth International Conference on Communication Systems and Network Technologies*, pp. 525–529, IEEE, Gwalior, India, April 2015.
- [12] B. Janarthanan, T. Santhanakrishnan, Real time metroplitan bus positioninsystem desing using GPS and GSM," in *Proceedings of the 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCCE)*, pp. 1–4, IEEE, Coimbatore, India, March 2014.
- [13] M. Kumbhar, M. Survase, P. Mastud, A. Salunke, and S. Sirdeshpande, "Real time web-based bus tracking system," *International Research Journal of Engineering and Technology*, vol. 3, no. 2, pp. 632–635, 2016.
- [14] J. Jinglin Li, J. Jie Gao, Y. Yu Yang, and H. Heran Wei, "Bus arrival time prediction based on mixed model," *China Communications*, vol. 14, no. 5, pp. 38–47, 2017.
- [15] V. E. Sathishkumar, W. A. Hatamleh, A. A. Alnuaim, M. Abdelhady, B. Venkatesh, and S. Santhoshkumar, "Secure dynamic group data sharing in semi-trusted third party cloud environment," *Arabian Journal for Science and Engineering*, pp. 1–9, 2021.
- [16] J. Ma, J. Theiler, and S. Perkins, "Accurate on-line support vector regression," *Neural Computation*, vol. 15, no. 11, pp. 2683–2703, 2003.
- [17] M. Mane and P. Suresh, V. D. Khairnar, Analysis of Bus Tracking System Using GPS on Smartphones," *IOSR Journal of Computer Engineering*, vol. 16, no. 2, 2014.
- [18] X.-G. Luo, H.-B. Zhang, Z.-L. Zhang, Y. Yu, and K. Li, "A new framework of intelligent public transportation system based on the internet of things," *IEEE Access*, vol. 7, pp. 55290–55304, 2019.
- [19] S. Chavhan, D. Gupta, B. N. Chandana, A. Khanna, and J. J. P. C. Rodrigues, "IoT-based context-aware intelligent public transport system in a metropolitan area," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6023–6034, 2020.
- [20] S. Chavhan, D. Gupta, R. K. Chidambaram, A. Khanna, and J. J. P. C. Rodrigues, "A novel emergent intelligence technique for public transport vehicle allocation problem in a dynamic transportation system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 5389–5402, 2021.
- [21] W. Treethidataphat, W. Pattara-Atikom, and S. Khaimook, "Bus arrival time prediction at any distance of bus route using deep neural network model," in *Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 988–992, IEEE, Yokohama, Japan, 2017 October.
- [22] J. Pan, X. Dai, X. Xu, and Y. Li, "A self-learning algorithm for predicting bus arrival time based on historical data model," in *Proceedings of the 2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems*, vol. 3, pp. 1112–1116, IEEE, Hangzhou, China, 2012 October.
- [23] P. He, G. Jiang, S.-K. Lam, and D. Tang, "Travel-time prediction of bus journey with multiple bus trips," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 11, pp. 4192–4205, 2019.
- [24] F. Pili, A. Olivo, and B. Barabino, "Evaluating alternative methods to estimate bus running times by archived automatic vehicle location data," *IET Intelligent Transport Systems*, vol. 13, no. 3, pp. 523–530, 2019.
- [25] E. Gayakwad, J. Prabhu, R. V. Anand, and M. S. Kumar, "Training time reduction in transfer learning for a similar dataset using deep learning," *Advances in Intelligent Systems and Computing*, pp. 359–367, 2021.
- [26] K. Dhyani, S. Bhachawat, J. Prabhu, and M. S. Kumar, "A novel survey on ubiquitous computing," *Data Intelligence and Cognitive Informatics*, pp. 109–123, 2022.
- [27] T. Pamuła and D. Pamuła, "Prediction of electric buses energy consumption from trip parameters using deep learning," *Energies*, vol. 15, no. 5, p. 1747, 2022.
- [28] Y. Ou, "AI for real-time bus travel time prediction in traffic congestion management," *Humanity Driven AI*, pp. 63–84, 2022.
- [29] N. Nagaraj, H. L. Gururaj, B. H. Swathi, and Y.-C. Hu, "Passenger flow prediction in bus transportation system using deep learning," *Multimedia Tools and Applications*, vol. 81, no. 9, pp. 12519–12542, 2022.
- [30] H. Abdelaty and M. Mohamed, "A framework for BEB energy prediction using low-resolution open-source data-driven model," *Transportation Research Part D: Transport and Environment*, vol. 103, Article ID 103170, 2022.
- [31] B. P. Ashwini, R. Sumathi, and H. S. Sudhira, "Bus travel time prediction: a comparative study of linear and non-linear machine learning models," *Journal of Physics: Conference Series*, vol. 2161, no. 1, Article ID 012053, 2022.
- [32] W. Lv, Y. Lv, Q. Ouyang, and Y. Ren, "A bus passenger flow prediction model fused with point-of-interest data based on extreme gradient boosting," *Applied Sciences*, vol. 12, no. 3, p. 940, 2022.

Retraction

Retracted: Human Resource Demand Prediction and Configuration Model Based on Grey Wolf Optimization and Recurrent Neural Network

Computational Intelligence and Neuroscience

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

Copyright © 2023 Computational Intelligence and Neuroscience. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.


The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] N. K. Rajagopal, M. Saini, R. Huerta-Soto et al., "Human Resource Demand Prediction and Configuration Model Based on Grey Wolf Optimization and Recurrent Neural Network," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 5613407, 11 pages, 2022.

Research Article

Human Resource Demand Prediction and Configuration Model Based on Grey Wolf Optimization and Recurrent Neural Network

Navaneetha Krishnan Rajagopal,¹ Mankeshva Saini,² Rosario Huerta-Soto,³
Rosa Vélchez-Vásquez,⁴ J. N. V. R. Swarup Kumar,⁵ Shashi Kant Gupta,⁶
and Sasikumar Perumal ⁷

¹Business Studies, University of Technology and Applied Sciences, Salalah, Oman

²Department of Management Studies, Government Engineering College Jhalawar, Jhalrapatan, Rajasthan, India

³Graduate School, Universidad Cesar Vallejo, Lima, Peru

⁴Faculty of Science, Universidad Nacional Santiago Antunez de Mayolo, Huaraz, Peru

⁵MIEEE, Department of Computer Science and Engineering, SR Gudlavalleru Engineering College, Gudlavalleru, India

⁶Computer Science Engineering, Integral University, Lucknow, UP, India

⁷Department of Computer Science, Wollo University, Kombolcha Institute of Technology, Kombolcha, Ethiopia Post Box No. 208

Correspondence should be addressed to Sasikumar Perumal; sasikumar@kiot.edu.et

Received 9 June 2022; Accepted 16 July 2022; Published 27 August 2022

Academic Editor: Zhongxu Hu

Copyright © 2022 Navaneetha Krishnan Rajagopal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Business development is dependent on a well-structured human resources (HR) system that maximizes the efficiency of an organization's human resources input and output. It is tough to provide adequate instructions for HR's unique task. In a time when the domestic labor market is still maturing, it is difficult for companies to make successful adjustments in HR structures to meet fluctuations in demand for human resources caused by shifting corporate strategies, operations, and size. Data on corporate human resources are often insufficient or inaccurate, which creates substantial nonlinearity and uncertainty when attempting to predict staffing needs, since human resource demand is influenced by numerous variables. The aim of this research is to predict the human resource demand using novel methods. Recurrent neural networks (RNNs) and grey wolf optimization (GWO) are used in this study to develop a new quantitative forecasting method for HR demand prediction. Initially, we collect the dataset and preprocess using normalization. The features are extracted using principal component analysis (PCA) and the proposed RNN with GWO effectively predicts the needs of HR. Moreover, organizations may be able to estimate personnel demand based on current circumstances, making forecasting more relevant and adaptive and enabling enterprises to accomplish their objectives via efficient human resource planning.

1. Introduction

Demand prediction for human resources (HR) is the practice of estimating the number and quality of personnel that will be required. For the forecast to be an accurate predictor, the annual budget and long-term company plan must be translated into activity levels for each function and department. Human resource demand forecasting is required to appropriately plan HR supply and demand. Implementing an enterprise development strategy may be aided by the development of an accurate human resource

demand forecasting model that is linked to the company's growth [1]. When it comes to predicting the needs of an organization's workforce, there are two components: demand and supply. The prediction of demand for HR is a precondition for the forecast of supply of HR. Human resource planning can only be done effectively if the demands of the company's future development are clearly defined because of the company's current situation and the supply and demand of HR. There will be a skill scarcity in the company if there are too many staff demand estimates, and this might impede the company's future growth as well [2].

The purpose of industrial development is to utilize natural resources and energy, as well as human resources, to aid industrial expansion by providing jobs and increasing exports. In today's competitive business world, organizations must continuously think about and adapt to the ever-changing environment to succeed. The company's products must be of the highest possible quality and inventive to meet the demands of a changing market. Competitiveness and long-term sustainability of an organization can only be achieved via the use of total quality management (TQM) and strategic human resource management (SHRM) [3]. Figure 1 depicts human resource planning.

Figure 1 explains that, to set out a strategy for human resource management, HR experts need a thorough awareness of their firm, as well as the ability to take into account different elements. Seven essential elements in the planning process may be used depending on the specifics of an enterprise: Determining the organization's goals is the first step, compiling a list of current employees is the second step, and the third step is to predict your human resources (HR) requirement; counting the number of skills gaps and their magnitude is the fourth step, making a plan of action is the fifth step, putting the strategy into action and integrating it with other aspects of your life is the sixth step, and monitoring, measuring, and providing feedback represent the final step.

Analyzing objectives is the initial step, inventory current human resources, forecast demand, estimate gaps, formulate plans, implement plans, monitor, control, and give feedback. Accurate forecasts of demand for HR may assist contemporary businesses to identify vacant or overstaffed positions and guide the logical distribution of HR. Because of this, it is important for the long-term viability of companies. It has resulted in various HR information systems developing decision-support functions and exploring ways to use existing HR data to enhance the allocation relationship between the internal staff and post requirements to solve a problem with HR allocation and provide scientific and rational support for the optimal positions [4]. A contemporary company's existence and resource development are dependent on its HR and its most valuable asset. To use HR effectively as well as the value and efficiency of HR is a key indicator of whether an enterprise's HR management has been successful or unsuccessful. Employee career development is becoming a more important aspect of HR growth in the workplace [5].

The remainder of the article is organized as follows: Section 2 provides a literature review and a problem statement. Proposed techniques are shown in Section 3. Section 4 contains the results. Section 5 is the discussion. Section 6 is the proposed work's conclusion.

2. Literature Review

According to study of [6], the notion of total quality (TQ) has gained a lot of traction in North America. Line management has always been concerned with total quality, which is based on the concept that firms may prosper by serving the demands of their consumers. Human and industrial

relations experts, on the other hand, have been advocating some of the ideas advanced by total quality converters for quite some time. In this context, their involvement in the implementation of a comprehensive quality strategy can only be beneficial. According to study of [7], TQM implementation was shown to be most influenced by "training and education," "incentive compensation," and "employee development" policies, according to research that examined the relationship between various HRM practices and TQM adoption. Human resource management adoption has the greatest influence on TQM procedures such as satisfaction of customers, statistical quality assurance, and cultural change and innovation. Also investigated were human resources management and total quality management as part of the research. In organizations that implemented HRM and TQM, "customer satisfaction" and "staff happiness" were strongly linked. According to study of [8], the HR scheduling model is based on the evaluation of HR data and the determination of the job matching score. Afterward, based on the job matching score, workers are scheduled. Grouping operations of neural networks are used as outputs in this study to increase neural network performance. An upgraded neural network is created once the data features are sorted and processed. To get the best possible results, for network configuration, we use a hierarchical paradigm. According to study of [9], HR are an organization's most important asset, and accurate demand forecasting is essential to making the most use of them. Predictive models are used to examine the company's human resource demands and define essential components of human resource allocation, starting with fundamental ideas of forecasting HR. This uses back-propagation neural networks (BPNN) and radial basis function neural networks (RBFNN). Two types of neural networks are used to anticipate current human resource needs based on past data. The outcomes of the predictions may be used by the company's managers to plan and allocate HR in a way that maximizes productivity. According to study of [10], the development of a country is heavily influenced by its HR. To improve the adaptability of specific-level strategy, forecasting demand for HR is done in both the commercial and governmental sectors. In addition, regular employment strengthens macroeconomic stability and fosters a sense of urgency for long-term prosperity. As outlined in this work, we use machine learning to predict the demand for human resources. According to study of [11], both a neural network-based dynamic learning prediction algorithm and an algorithm for optimizing resource allocation are proposed in this article; HR may be organized around just two shifts thanks to these two algorithms, which lessen the unpredictability of ship arrivals. The opposite is also true: operators can be optimally distributed throughout the day, taking into consideration real demand and the terminal's operations. In addition, because these algorithms are based on universal variables, they may be used at any transshipment port. According to study of [12], HR for health planning should be in sync with health scheme requirements for an efficient health system. To support HRH programs and policies, it is necessary to create strategies for quantifying the requirements and supply of health workers.

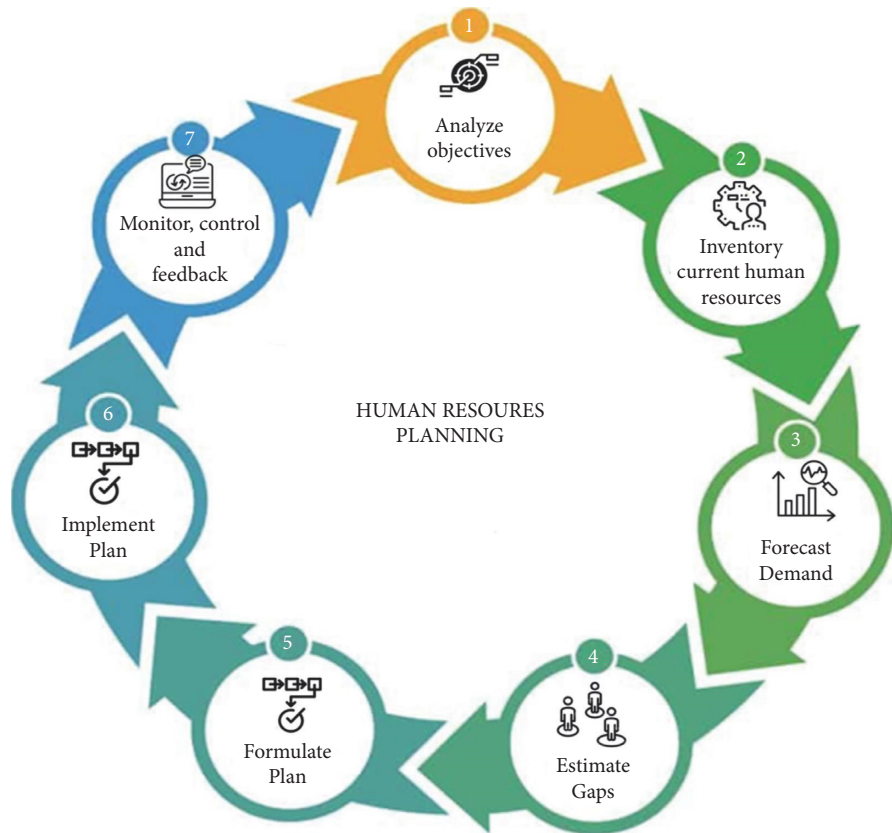


FIGURE 1: Human resource planning.

Secondary data on service use and population projections, as well as expert opinions, were the primary sources of information for this investigation. Using the health demand technique, the HRH requirements were estimated based on the anticipated service utilizations. The staffing standard and productivity were used to transform into HRH requirements. According to study of [13], multiskilled HR in R&D projects may be allocated using an optimization model that considers each worker's unique knowledge, experience, and ability. In this approach, three key characteristics of HR are taken into consideration: the various skill levels, the learning process, and the social interactions that occur within working groups. There are two approaches to resolving the multiobjective problem: The optimal Pareto frontier is first explored to find a collection of nondominated solutions, and then, armed with new knowledge, the ELECTRE III approach is used to find the best compromise between the various goals. Each solution's uncertainty is represented by fuzzy numbers, which are then used to calculate the ELECTRE III's threshold values. The weights of the objectives are then calculated based on the relative importance of each objective to the others. According to study of [14], a multiagent approach for allocating HR in software projects that are distributed across many time zones is introduced by work. When a project needs HR, this mechanism takes into account the context of the project participants, the requirements of the activities, and the interpersonal interaction between those people. Contextual information provided

by the participants includes things like culture, language, proximity in time, prior knowledge, intelligence collection, and reasoning, and allocation of HR falls within the purview of this system. According to study of [15], the design and implementation of flexible information systems rely heavily on knowledge of how businesses work. There are a variety of procedures that companies go through daily. To carry them out, a series of interconnected events and actions or tasks must be completed. Additionally, it incorporates key decision points and the individuals involved in carrying out the process to provide a final deliverable that comprises one or more outputs. According to study of [16], allocating resources at design time and runtime is a common task for business process management systems to undertake. This study aims to fill up the knowledge gap. User preference models for semantic web services have been proven and versatile; therefore we provide a method to define resource preferences. In addition, we show the approach's practicality by implementing one. According to study of [17], an organization's operations are orchestrated with the help of business process management systems. These systems use information about resources and activities to determine how to distribute resources to accomplish a given task. It is commonly accepted that resource allocation may be improved by taking into account the characteristics of the resources that are being considered. The Fleishman taxonomy may be used to identify activities and HR. To allocate resources throughout the process runtime, these specs are

employed. We demonstrate how a business process management system may implement the ability-based allocation of resources and assess the technique in a realistic situation. According to study of [18], one of the most important aspects of an organization's viewpoint is allocating the most appropriate resource to carry out the operations of a business process. The business processes may benefit from increased efficiency and effectiveness if the resources responsible for carrying out the activities are better selected. On behalf of enterprises, we have defined and categorized the most important criteria for resource allocation methodologies. Criteria about HR were the primary focus of our investigation. It is our aim that the proposed classification would aid those in charge of process-oriented systems in discovering the sort of information needed to assess assets. As a result of this categorization, additional resource-related information may be captured and integrated into BPMS systems, which might improve the present support for the organizational viewpoint. Additional criteria for evaluation will be added, and we intend to investigate the effects of those factors on resource allocation and codify the criteria for resource allocation identified in taxonomy of resource allocation criteria. According to studies of [19, 20], human resource allocation is further complicated by the presence of team fault lines, which are detailed in this work. Using the information value, we first examine resource characteristics from a demographic and business process viewpoint before selecting essential qualities and assigning a weight to them. This is followed by qualitative and quantitative analysis of team fault lines based on the clustering results of HR. The base and ensemble performance prediction model is built using a multilayer perception. Subsequently, the allocation model and flow are developed. In a real-world scenario, the rationality and efficacy of our human resource allocation approach employing team fault lines were examined, with findings showing that our method can effectively distribute HR and optimize business processes. According to study of [21], an on-the-fly allocation of HR using Naïve Bayes is proposed in this paper. Resource allocation plans are said to be updated and performed "on the fly" in this context, meaning that current human resource performance is taken into account while they are being implemented. Our research shows that the suggested methodology takes less time overall to complete than existing methods of allocating resources. The researchers hypothesized, using a numerous constituency perspective of the HR function, that organizational financial investment in their HR functions will have an impact on labor productivity and that this relationship will be moderated by the presence of professional HR staff and the adoption of high performance work systems [22, 23]. Selection, training, working conditions, and assessment were included as independent variables in this study's analysis of HR planning while job satisfaction as a proxy for organizational performance was used as the dependent variable. A self-rated questionnaire has been issued to the organization's top level, medium level, and lower level managers who have read current and prior extensive literature on the importance of human resource practice in organizations [24].

2.1. Problem Statement. Nonlinearity and unpredictability in each component's relationship to the demand for HR are considerable, as are the incompleteness and inaccuracy of corporate human resource data. However, the workforce planning process reveals that many organizations are unsatisfied with their ability to convert company strategy into the particular numbers of personnel needed to fulfill business objectives. It was commonly thought that demand forecasting, or figuring out how many employees are needed, was one of the most difficult aspects of managing labor shortages. Disaster relief organizations have several challenges, including limited human and financial resources and unpredictability in disaster assistance environments. Despite this, no single demand forecasting model has been established to address the aforementioned issue.

3. Proposed Methodology

In this phase, we examine the human resource demand prediction and configuration model based on grey wolf optimization and recurrent neural network. Figure 2 depicts the overall methodology used.

In Figure 2, the data is collected, and the data can be preprocessed using normalization. Principal component analysis (PCA) is used for feature extraction. A recurrent neural network is used for prediction and grey wolf optimization is used for human resource demand prediction.

3.1. Data Collection. The 9,855 unique employee users in our dataset represent nearly 15% of the workforce. Over a year and a half, we gathered all 15,200 communications with clear reply links. We have cleaned up the message content by using a stemming step; after normalization phase, we were able to extract 4,384 unique terms from the text of all communications. Additionally, we have collected information on the companies in which the employees who have signed up for our corporate micro blogging platform have been employed (a subset of the total business hierarchy). We create a profile for each person that provides data on their present and prior roles within the organization, as well as a timeline of their previous work history, projects they worked on, and so on [25].

3.2. Data Preprocessing Using Normalization. Typically, healthcare databases are made up of a range of heterogeneous data sources, and the data extracted from them are different, partial, and redundant, all of which have a significant impact on the final mining outcome. As a result, healthcare data must be preprocessed to guarantee that they are accurate, full, and consistent and have privacy protection. Normalization is a preprocessing approach in which the data are scaled or altered to ensure that each feature contributes equally to the total. It is possible to construct a new range from an existing one using the normalization procedure. Predictions or forecasts based on this information may be very valuable. Each feature contributes the same amount of data whether the raw data are rescaled or transformed. Outliers and dominant features, two

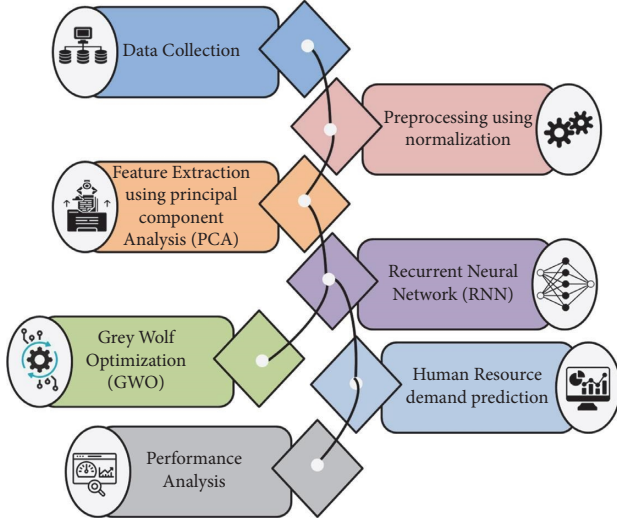


FIGURE 2: Overall methodology used.

significant data problems that impede machine learning algorithms, are addressed here. Based on statistical measurements from raw (unnormalized) data, several ways for normalizing data within a specified range have been devised. We normalized our data using the Min-Max and Z-score methods. According to the way raw data statistical characteristics are used to normalize the data, these techniques are categorized.

Min-Max normalization is a technique for converting data linearly at the start of the range. Using this method, the relationship between different pieces of information is preserved. Predefined borders with predefined boundaries are a crucial strategy that can correctly fit information.

Following this approach to normalization,

$$O' = \left(\frac{O - \text{min value of } O}{\text{max value of } O - \text{min value of } O} \right) * (R - I) + I. \quad (1)$$

Min-Max data is included in O , and one of the boundaries is $[I, R]$.

The range of the real data is denoted by O , while the mapped data is denoted by O' .

In the Z-score normalization procedure, the mean and SD of the data are used to obtain a normalized value from unstructured information. As can be seen in (2), the unstructured data may be normalized using the Z-score variable.

$$f'_i = \frac{f_i - \bar{Z}}{\text{std}(Z)}, \quad (2)$$

where f'_i shows the standardised Z-score values and f_i shows which l^{th} column's row Y the value is in.

$$\text{std}(Z) = \sqrt{\frac{1}{(M-1)} \sum_{l=1}^m (f_l - \bar{Z})^2}, \quad (3)$$

$$\bar{Z} = \frac{1}{m} \sum_{l=1}^m f_l \text{ or mean value.}$$

In this example, the variables or columns that begin with "P" are found in each of the rows $D, E, F,$ and G through H . Z-score approach may be used in each row since every value in a row is equal, resulting in zero standard deviation, and every value in that row is set at 0 to generate standard data. Min-Max normalization is similar in that it shows the range of values between 0 and 1, as is the Z-score.

Scaling by decimal points is the method that allows for the range of -1 to 1 . In line with this strategy,

$$f^l = \frac{g}{10^s}. \quad (4)$$

Here, f^l indicates the values scaled, g represents the value range, and s denotes the smallest integer $\text{Max}(|f^l|) < 1$.

3.3. Feature Extraction Using Principal Component Analysis (PCA). In this study, we extract features using principal component analysis (PCA), which yields encouraging results.

3.3.1. Principal Component Analysis (PCA). PCA is a technique for reducing the number of dimensions in a dataset. When a high-dimensional dataset is reduced to a smaller dimension, PCA transforms it into a least-squares projection. Datasets can be reduced in dimensionality by discovering a new collection of variables smaller than the original set that represents the significant primary variability in the data. Most of the sample's information remains intact as PCA extracts essential details from complex datasets. It is a straightforward, nonparametric approach to reducing dimensions. As a result, data compression and categorization can benefit from it. In a wide range of industries, principal component analysis has been utilized. Pattern recognition and other data reduction techniques are good examples of picture processing and compression. Computer-Aided Discrepancy Analysis is a technique for finding the direction of the highest variation within a given input space and calculating the covariance matrix's principal component.

The following is the algebraic definition of PCA.

Calculate the covariance of Y and the mean of Y for data matrix Y .

$$\mu := F\{Z\}, \quad (5)$$

Then compute the correlation coefficient.

$$S = \text{COV}(Y) = F\{(Y - \mu)\}(\mu)(Y - \mu)^N. \quad (6)$$

In order to count the eigenvectors $f_1, f_2, \dots, f_O, j = 1, 2, \dots, O$ and eigenvalues λ_j of the covariance S eigenvalues, sort the values in decreasing order.

Solving the equation for S , the covariance

$$|\lambda J - S| = 0. \quad (7)$$

Get the eigenvalues by decomposing using SVD.

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q \geq 0. \quad (8)$$

With the corresponding eigenvectors, f_j ($j = 1, 2, \dots, q$).

Counting the first N eigenvalues is used to pick λ_j to retrieve the principal components.

$$\frac{\sum_{j=1}^N \lambda_j}{\sum_{k=1}^O \lambda_j} \quad (9)$$

The first N eigenvalue that increased the cumulative percentage by 85 percent is selected as the major component.

The data are projected into a smaller subspace with fewer dimensions.

$$Q = X^U Y. \quad (10)$$

We may minimize the number of variables or dimensions from o to N ($N \ll o$) by utilizing the first N respective eigenvectors.

The principal component analysis aims to identify linear combinations of factors that best describe the data. We are experimenting with PCA as a tool for extracting features and shrinking dimensions. We may now experiment with a wide range of various classification or grouping methods based on the data we have acquired.

3.3.2. Recurrent Neural Network (RNN). We look at probability distributions defined over discrete sample space, with a single configuration consisting of $\alpha \equiv (\alpha_1, \alpha_2, \dots, \alpha_O)$ list of O variables α_o and $\alpha_o \in \{0, \dots, e_w - 1\}$. The input dimension e_w denotes the number of potential values for each variable α_o . Figure 3 shows that.

In circumstances when variables α_o have high correlations, one of the most important tasks in machine learning is to infer probability distributions from a collection of empirical data. We utilize the product rule for probabilities to describe the likelihood of a configuration α as $Q(\alpha) \equiv Q(\alpha_1, \alpha_2, \dots, \alpha_O)$.

$$Q(\alpha) = Q(\alpha_1)Q(\alpha_2|\alpha_1) \dots Q(\alpha_O|\alpha_{O-1}, \dots, \alpha_2, \alpha_1), \quad (11)$$

where $Q(\alpha_j|\alpha_{j-1}, \dots, \alpha_2, \alpha_1) \equiv Q(\alpha_j|\alpha_{<j})$ is the conditional distribution of α_j given a configuration of all α_j with $k < j$.

RNNs are a kind of correlated probability distribution of the form (11), in which $Q(\alpha)$ is defined by the conditionals $Q(\alpha_j|\alpha_{<j})$. A recurrent cell, which has appeared in many forms in the past, is the basic building unit of an RNN.

A recurrent cell is a nonlinear function that transfers the direct sum (or concatenation) of an incoming hidden vector i_{o-1} of dimension d_i and an input vector α_{o-1} to output is hidden vector i_o of dimension d_i in the simplest form possible.

$$i_o = g(X[i_{o-1}; \alpha_{o-1}] + c). \quad (12)$$

A nonlinear activation function is g .

The weight matrix $X \in \mathbb{R}^{e_i \times (e_i + e_o)}$, the bias vector $c \in \mathbb{R}^{e_i}$, and the states $i_o \alpha_o$ that initiate the recursion are the parameters of this basic RNN ("vanilla" RNN). We set i_0 and α_0 to constant values in this study. Vector α_n encodes the input α_n in a single pass. The whole probability $Q(\alpha)$ is computed by successively calculating the conditionals, beginning with (α_1) , as follows:

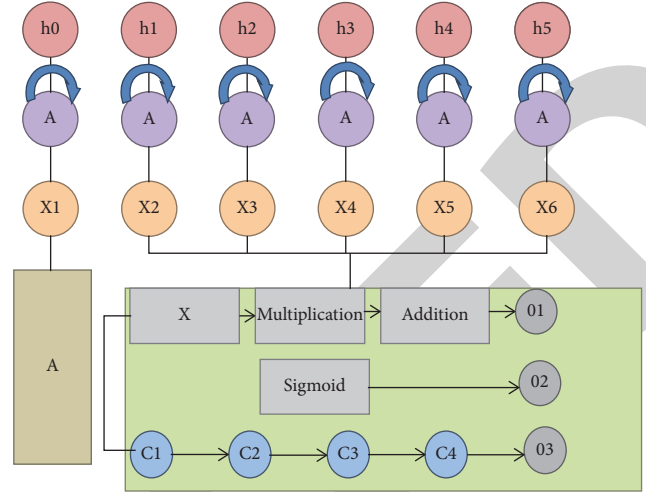


FIGURE 3: Framework of RNN.

$$p(\alpha_o|\alpha_{o-1}, \dots, \alpha_1) = z_o \cdot \alpha_o, \quad (13)$$

where the right-hand side has the standard scalar product of vectors and

$$z_o \equiv T(Vi_o + d). \quad (14)$$

$V \in \mathbb{R}^{e_w \times e_i}$ and $d \in \mathbb{R}^{e_w}$ are the weights and biases of a Softmax layer, respectively, while T is the Softmax activation function.

$$T(w_o) = \frac{\exp(w_o)}{\sum_j \exp(w_j)}, \quad (15)$$

In (13) $z_o = (z_o^1, \dots, z_o^{e_w})$ is an e_w -component vector of positive, real values that add up to 1.

$$z_{o1} = 1. \quad (16)$$

Thus, a probability distribution over states α_o is formed. The entire probability $Q(\alpha)$ is given by

$$Q(\alpha) = \prod_{n=1}^O z_o \cdot \alpha_o. \quad (17)$$

Note that $Q(\alpha)$ is already properly normalized to unity such that

$$Q(\alpha)_1 = 1. \quad (18)$$

The hidden vector i_o encodes information about prior spin configurations $\alpha_{<o}$, as shown in (12) and (13). History $\alpha_{<o} \varepsilon$ is important for predicting the probability of subsequent α_o for correlated probabilities. The RNN is capable of simulating tightly linked distributions by transmitting hidden states in (13) between sites. The dimension of the concealed state will be referred to as the number of memory units i_o . Figure 3 shows the framework of RNN.

3.4. Grey Wolf Optimization (GWO). The grey wolf optimization (GWO) algorithm is a new bio-inspired optimization approach. The main goal of the GWO method is

to find the best solution for a given issue utilizing a population of search agents. The social dominance hierarchy that creates the candidate solution in each iteration of optimization distinguishes the GWO algorithm from other optimization algorithms. Tracking, surrounding, and assaulting the target are the three processes in the hunting mechanism. Thus, GWO stands for the grey wolf's mathematical hunting approach, which is utilized to tackle complex optimization problems. As a result, the best solution to a problem has been deemed a victim.

The three upper levels' movement represents the victim being encircled by grey wolves, as stated by the following formula:

$$\vec{E} = \vec{D} \cdot \overrightarrow{Y_q(u)} - \vec{Y}(u). \quad (19)$$

Y_q indicates the prey position vector, Y represents the grey wolf location, and D is the coefficient vector. Using the following equation, the result of vector E is used to shift a specific element closer to or away from the region where the optimal solution, which symbolizes the prey, is placed.

$$\overrightarrow{y(t+1)} = \left| \overrightarrow{y_r(t) - \vec{c} \cdot \vec{d}} \right|, \text{ with } \vec{c} = 2\vec{c} \cdot m_1 - \vec{c}, \quad (20)$$

where s_1 is chosen at random from the range $[0, 1]$, and over a predetermined number of repetitions, a is decreased from 2 to 0. If $|B|$ is greater than one, this corresponds to exploitation behavior and replicates prey attack behavior. If $|B| > 1$, the wolf spacing from the victim is imitated. $[-2, 2]$ are the recommended values for A . Using the following mathematical equations, three higher levels, a , b , and c , will be calculated.

$$\begin{aligned} \vec{E}_a &= \left| \vec{D}_1 \cdot \vec{Y}_a - \vec{Y} \right| \text{ with } \vec{Y}_1 = \vec{Y}_a - \vec{Y}_a \cdot (\vec{E}_a), \\ \vec{E}_b &= \left| \vec{D}_2 \cdot \vec{Y}_b - \vec{Y} \right| \text{ with } \vec{Y}_2 = \vec{Y}_b - \vec{Y}_b \cdot (\vec{E}_b), \\ \vec{E}_c &= \left| \vec{D}_3 \cdot \vec{Y}_c - \vec{Y} \right| \text{ with } \vec{Y}_3 = \vec{Y}_c - \vec{Y}_c \cdot (\vec{E}_c). \end{aligned} \quad (21)$$

Assume that a , b , and c have enough information about the likely whereabouts of the victim to mathematically imitate the grey wolf's hunting method. Furthermore, the top three best solutions are preserved, forcing the other agents to update their positions following the best agents a , b , and c . The pseudocode of the GWO is given in Algorithm 1, and this behavior is mathematically represented by the following statement.

$$\vec{Y}(u+1) = \frac{\vec{y}_1 + \vec{y}_2 + \vec{y}_3}{3}. \quad (22)$$

4. Result

In this phase, we examine the human resource demand prediction and configuration model based on grey wolf optimization and recurrent neural network. The parameters are in-demand analytics skills, human resource satisfaction index (HRSI), prediction rate, and error rate. The existing methods are convolutional neural network (CNN [26]), double cycle neural network (DCNN [27]), whale optimization (WO [8]), and particle swarm optimization (PSO [28]).

4.1. In-Demand Analytics Skills. In demand analytics skills, we assessed the employee experience, people analysis, internal recruiting, and multigenerational workforce. The in-demand analytics skills are depicted in Figure 4.

Figure 4 explains employee experience with a score of 94 percent, people analytics with a score of 85 percent, internal recruitment with a score of 82 percent, and a multigenerational workforce with a score of 74 percent. While comparing the employee experience, people analytics, and internal recruiting with a multigenerational workforce, it is shown that the multigenerational workforce is lower than the others.

4.2. Human Resource Satisfaction Index (HRSI). In HRSI, we assessed the employer image, employee expectations, perceived HR service quality, value perceived by the employee, employee satisfaction, employee loyalty, and HRSI. The human resource satisfaction index is depicted in Figure 5.

Figure 5 shows employer image with a score of 34.94 percent, employer expectations with a score of 51.62 percent, perceived HR service quality with a score of 71.35 percent, value perceived by the employee with a score of 70.43 percent, employee satisfaction with a score of 54.45 percent, employee loyalty with a score of 42.23 percent, and HRSI with a score of 54.17 percent. While comparing employer image, employer expectations, perceived HR service quality, employee satisfaction, employee loyalty, and HRSI, it is shown that perceived HR service is higher than the others.

4.3. Prediction Rate. In other words, if an early warning system can accurately predict a need for HR, then it has a high predictive value. Figure 6 represents the prediction rate.

In Figure 6, we evaluate the convolutional neural network with a prediction rate of 73 percent, the double cycle neural network with a prediction rate of 63 percent, the whale optimization with a prediction rate of 85 percent, and the particle swarm optimization with a prediction rate of 58 percent, and we propose RNN + GWO with a prediction rate of 95 percent. The results of the comparisons reveal that the suggested approach is superior to each of the four methods that already exist.

```

Set the grey wolf population  $Y_j$  ( $j = 1, 2, \dots, o$ )
Set up  $b$ ,  $B$ , and  $d$ .
Compute the search agent's fitness
The most effective search agent =  $Y_a$ 
An agent with the second-best search =  $Y_b$ 
the third most effective search agent =  $Y_c$ 
while ( $u < \text{maximumno of iteration}$ )
    for every search agent
        Change the current search agent's position.
    end for
    Improve  $a$ ,  $A$ , and  $c$ 
    Adjust the current position of the search agent
    Improve  $X_\alpha$ ,  $X_\beta$ , and  $X_\delta$ 
     $t = t + 1$ 
end while
return  $X_a$ 

```

ALGORITHM 1: Grey wolf optimization (GWO).

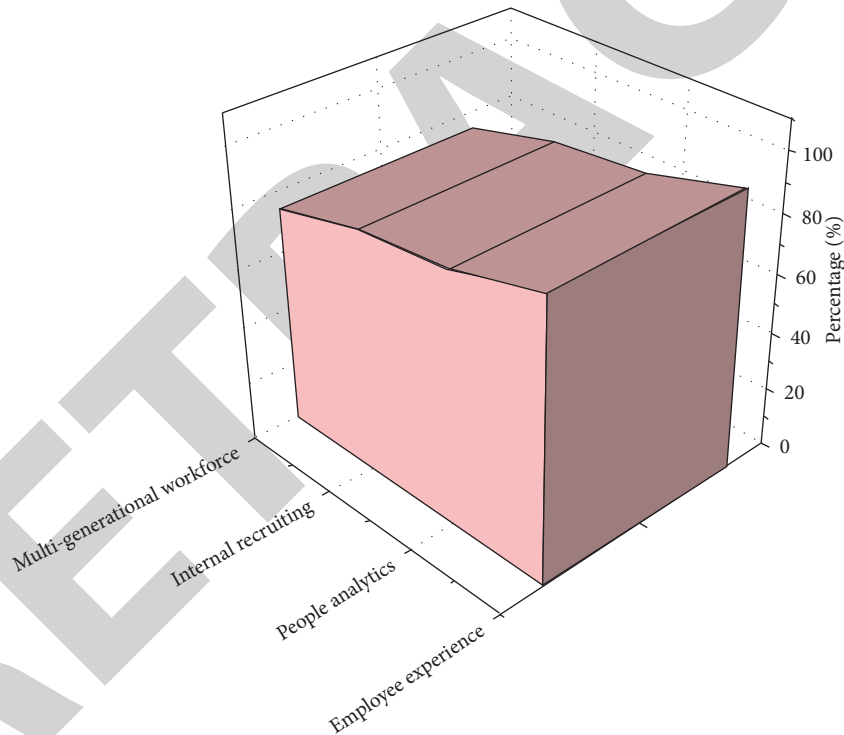


FIGURE 4: In-demand analytics skills.

4.4. Error Rate. It is the percentage of incorrect units of data transferred in comparison to the overall number of units of data. Figure 7 shows the error rate.

In Figure 7, we evaluate the convolution neural network with an error rate of 93 percent, the double cycle neural network with an error rate of 85 percent, the whale optimization with an error rate of 80 percent, and the particle swarm optimization with an error rate of 75 percent, and we propose RNN+GWO with an error rate of 50 percent.

The results of the comparisons demonstrate that the suggested approach is inferior to each of the four other strategies already in use.

5. Discussion

In CNN (existing), this model presents several problems, the most significant of which are overfitting, inflating gradients, and class unbalances. The effectiveness of the

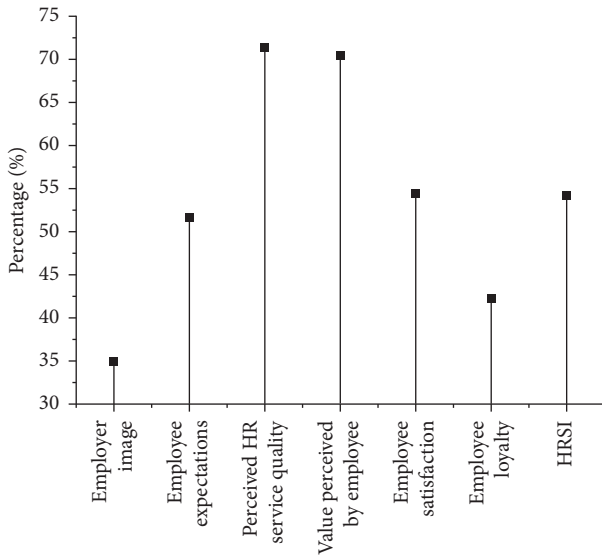


FIGURE 5: Human resource satisfaction index (HRSI).

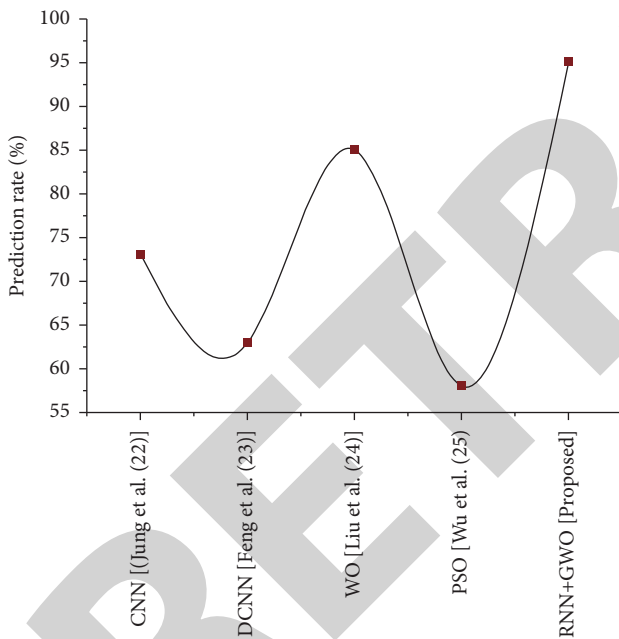


FIGURE 6: Comparative analysis of prediction rate for suggested and traditional methods.

model may suffer as a result of these concerns. Negative aspects of the DCNN (existing) include the following: the lifetime of the network is uncertain; the working of the network is not described, and it is difficult to demonstrate the issue to the network. The whale optimization (existing) suffers from some flaws, the most notable of which are its sluggish convergence, poor solution accuracy, and the ease with which it might fall into the local optimum solution. The particle swarm optimization (PSO) technique has several drawbacks, the most notable of which are that it is simple to become stuck in a local optimum in a high-dimensional space and that

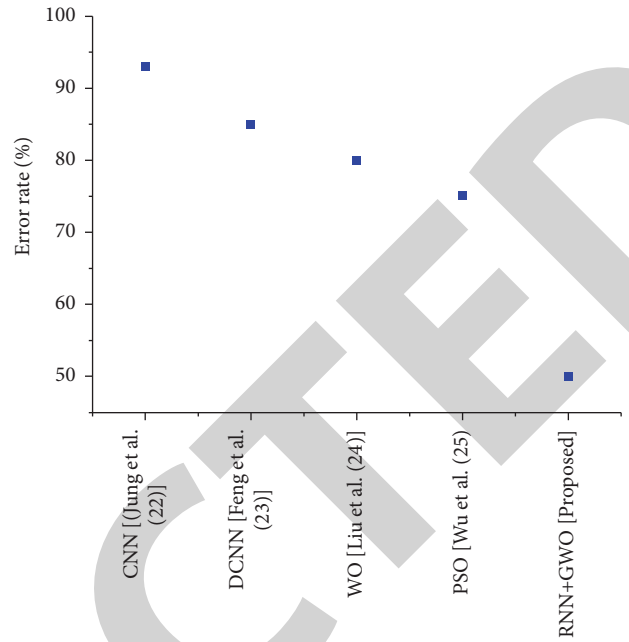


FIGURE 7: Comparative analysis of error rate for suggested and traditional methods.

it has a slow convergence rate throughout the repeated process.

6. Conclusion

The key to an enterprise’s long-term success is its ability to effectively use its HR. There will be a great amount of data on HR created inside the firm as the company continues to grow and expand. Data on corporate HR are often inadequate or inaccurate since the demand for HR is impacted by so many factors. As a result, there is a high degree of nonlinearity between various components and demand for HR. Human resource demand may be forecasted using a recurrent neural network (RNN) and grey wolf optimization (GWO) which is a revolutionary quantitative forecasting approach of tremendous theoretical significance. The first step is to get the data and normalize it. Principal component analysis (PCA) is used to identify the characteristics, and the suggested RNN with GWO can accurately estimate human resource requirements. To make forecasting more relevant, flexible, and accurate, the human resource demand prediction model was developed and it might help organizations better manage their HR to meet their priorities. El Health employs IoT sensors in particular to monitor employee demand, which is analyzed as a time-series data and utilized to forecast future need.

Data Availability

No data were used in this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Y. Zhang, L. I. Feng, and L. I. U. Bo, "Research on the forecasting model of total human resource demand of large central enterprise groups based on the CD production function," *DEStech Transactions on Computer Science and Engineering*, vol. 4, pp. 77–79, 2017.
- [2] J. Zheng and R. Ma, "Analysis of enterprise human resources demand forecast model based on SOM neural network," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 6596548, 10 pages, 2021.
- [3] L. Hataani and S. Mahrani, "Strategic human resource management practices: mediator of total quality management and competitiveness (a study on small and medium enterprises in kendari southeast sulawesi)," *International Journal of Business and Management Invention*, vol. 2, no. 1, pp. 8–20, 2013.
- [4] S. Cao and Q. Du, "Forecasting of human resources needs to be based on BP neural network," *Shandong Technology University Journal*, vol. 22, no. 5, pp. 26–29, 2008.
- [5] L. Wang, "Analysis of modern enterprise employee career development mode innovation," vol. 105, pp. 396–400, in *Proceedings of the In 2015 3rd International Conference on Education, Management, Arts, Economics and Social Science*, vol. 105, Atlantis Press, Amsterdam, Netherlands, December 2015.
- [6] M. Izvercian, A. Radu, L. Ivascu, and B. O. Ardelean, "The impact of human resources and total quality management on the enterprise," *Procedia - Social and Behavioral Sciences*, vol. 124, pp. 27–33, 2014.
- [7] C. C. Yang, "The impact of human resource management practices on the implementation of total quality management," *The TQM Magazine*, vol. 18, no. 2, pp. 162–173, 2006.
- [8] Y. Liu, W. Zhang, Q. Zhang, and M. Norouzi, "An Optimized Human Resource Management Model for Cloud-Edge Computing in the Internet of Things," *Cluster Computing*, vol. 25, pp. 2527–2539, 2021.
- [9] S. Yuan, Q. Qi, E. Dai, and Y. Liang, "Human Resource Planning and Configuration Based on Machine Learning," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 3605722, 6 pages, 2022.
- [10] K. S. Nguyen, H. D. Hung, V. T. Tran, and T. A. Le, "An approach to human resource demand forecasting based on machine learning techniques," in *Research in Intelligent and Computing in Engineering*, pp. 389–396, Springer, Singapore, 2021.
- [11] G. Fancello, C. Pani, M. Pisano, P. Serra, P. Zuddas, and P. Fadda, "Prediction of arrival times and human resources allocation for container terminal," *Maritime Economics & Logistics*, vol. 13, no. 2, pp. 142–173, 2011.
- [12] N. Pagaiya, P. Phanthunane, A. Bamrung, T. Noree, and K. Kongweerakul, "Forecasting imbalances of human resources for health in the Thailand health service system: application of a health demand method," *Human Resources for Health*, vol. 17, no. 1, pp. 4–12, 2019.
- [13] L. O. Teixeira and E. H. Huzita, "DiSEN-AlocaHR: a multi-agent mechanism for human resources allocation in a distributed software development environment," in *Proceedings of the Distributed Computing and Artificial Intelligence, 11th International Conference*, pp. 227–234, Springer, Spain, June 2014.
- [14] M. Dumas, M. L. Rosa, J. Mendling, and H. A. Reijers, *Fundamentals of Business Process Management* vol. 1, p. 2, Springer, Berlin, Germany, 2013.
- [15] H. Guo, R. Brown, and R. Rasmussen, "Human resource behavior simulation in business processes," *Information systems development, reflections, challenges, and new directions, proceedings of ISD 2011*, Heriot-Watt University, vol. 56, no. 2, , pp. 376–405, Edinburgh, Scotland, 2013.
- [16] C. Cabanillas, J. M. García, M. Resinas, D. Ruiz, J. Mendling, and A. Ruiz-Cortés, "Priority-based human resource allocation in business processes," vol. 274, pp. 374–388, in *Proceedings of the Service-Oriented Computing*, vol. 274, Springer, Berlin, Germany, December 2013.
- [17] J. Erasmus, I. Vanderfeesten, K. Traganos, A. Jie A Looi, P. Kleingeld, and P. Grefen, "A method to enable ability-based human resource allocation in business process management systems," vol. 335, pp. 37–52, in *Proceedings of the Lecture Notes in Business Information Processing*, vol. 335, Springer, Cham, October 2018.
- [18] M. Arias, J. Munoz-Gama, and M. Sepúlveda, "Towards a taxonomy of human resource allocation criteria," vol. 308, pp. 475–483, in *Proceedings of the International Conference on Business Process Management*, vol. 308, pp. 475–483, Springer, Cham, September 2017.
- [19] W. Zhao, S. Pu, and D. Jiang, "A human resource allocation method for business processes using team fault-lines," *Applied Intelligence*, vol. 50, no. 9, pp. 2887–2900, 2020.
- [20] Z. Hu, Y. Zhang, Y. Xing, Y. Zhao, D. Cao, and C. Lv, "Toward human-centered automated driving: a novel spatial-temporal vision transformer-enabled head tracker," *IEEE Vehicular Technology Magazine*, 2022.
- [21] A. Wibisono, A. S. Nisafani, H. Bae, and Y. J. Park, "On-the-fly performance-aware human resource allocation in the business process management systems environment using naïve Bayes," vol. 219, pp. 70–80, in *Proceedings of the Lecture Notes in Business Information Processing*, vol. 219, Springer, Busan, Republic of Korea, June 2015.
- [22] M. Subramony, J. P. Guthrie, and J. Dooney, "Investing in HR? Human resource function investments and labor productivity in US organizations," *International Journal of Human Resource Management*, vol. 32, no. 2, pp. 307–330, 2021.
- [23] Z. Hu, Y. Xing, W. Gu, D. Cao, and C. Lv, "Driver anomaly quantification for intelligent vehicles: a contrastive learning approach with representation clustering," *IEEE Transactions on Intelligent Vehicles*, p. 1, 2022.
- [24] S. Muhammad, "Practice of human resource planning in organizations: a study based on organizational performance," *KASBIT Business Journal*, vol. 15, no. 2, pp. 102–114, 2022.
- [25] H. Wu, C. Chelmis, V. Sorathia, Y. Zhang, O. P. Patri, and V. K. Prasanna, "Enriching employee ontology for enterprises with knowledge discovery from social networks," in *Proceedings of the 2013 IEEE/AiCM International*

Research Article

A Massage Area Positioning Algorithm for Intelligent Massage System

Liran Zhou ^{1,2}, Zhiquan Feng ^{1,2}, Zeyuan Cai^{1,2}, Xiaohui Yang ^{1,2,3}, Changsheng Ai,⁴ and Haiyan Shao⁴

¹School of Information Science and Engineering, University of Jinan, Jinan 250022, China

²Shandong Provincial Key Laboratory of Network Based Intelligent Computing, Jinan 250022, China

³State Key Laboratory of High-end Server & Storage Technology, Jinan 250002, China

⁴School of Mechanical Engineering, University of Jinan, Jinan 250022, China

Correspondence should be addressed to Zhiquan Feng; ise_fengzq@ujn.edu.cn

Received 11 May 2022; Accepted 13 July 2022; Published 4 August 2022

Academic Editor: Zhongxu Hu

Copyright © 2022 Liran Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A growing number of studies have been conducted over the past few years on the positioning of daily massage robots. However, most methods used for research have low interactivity, and a systematic method should be designed for accurate and intelligent positioning, thus compromising usability and user experience. In this study, a massage positioning algorithm with online learning capabilities is presented. The algorithm has the following main innovations: (1) autonomous massage localization can be achieved by gaining insights into natural human-machine interaction behavior and (2) online learning of user massage habits can be achieved by integrating recursive Bayesian ideas. As revealed by the experimental results, combining natural human-computer interaction and online learning with massage positioning is capable of helping people get rid of positioning aids, reducing their psychological and cognitive load, and achieving a more desirable positioning effect. Furthermore, the results of the analysis of user evaluations further verify the effectiveness of the algorithm.

1. Introduction

As health care has aroused the rising attention of people, it has become a trend to integrate intelligent massage robots into people's daily lives to help them achieve massage and relaxation anytime and anywhere. Massage positioning is recognized as the first task of an intelligent massage robot, and numerous existing research results have been achieved relating to massage positioning.

Three main types of massage systems have been used on the market over the past few years. The first one is the larger massage chair, which is the most common and requires the operation of a remote control or control panel to give the massage positioning instructions. This positioning mode of operation, however, is not friendly to the elderly and has a rigid division of massage areas. The second type refers to a small regional massager (e.g., a neck massager). The above type of massager is focused and can only be applied in one

area. The third type refers to the professional type massage robot arm, which is primarily employed in professional massage hospitals or massage shops. The user should lie in a fixed position after completing the diagnosis, and the massage position is set by the professional massage practitioner independently. This massage robot is characterized by strong professionalism and high precision, whereas the system requires external support staff for massage positioning. Accordingly, this positioning model runs counter to the goal of integrating massage robots into daily lives. In brief, current massage positioning systems are always dependent on external conditions (e.g., massage manipulation boards and professional doctors) and are not intelligent. Accordingly, understanding the natural way in which people express themselves in the massage area can help the user disengage from the above external dependencies.

It is generally known that the ability to express intentions through physical movements is the most basic and common

ability among people [1]. It has been found that when people express something or an area, they usually use gestures or words to direct the attention of others to the target, thus enabling them to gain insights into their intentions [2]. In the case of gestures, people usually use pointing actions to express the target when they are far away from it, that is, indicating behavior [3], which is commonly achieved by extending the index finger and flexing the rest of the fingers [4]. If the target is close, people usually express it in a contact manner, for example, by holding or touching it directly with their hand. Likewise, the above results can be applied to people's representations of the massage area. People typically use static pointing expressions for areas of the body that are distant from the hand. For closer areas, on the other hand, direct contact with the fingertips of the index finger is generally used for expression. On that basis, pointing gestures have become one of the most natural ways of expressing the massage area, and verbal expressions are also one of the most natural ways. Thus, a correct understanding of the above natural expressions can help people disengage from the operating board and reduce memory and operational load. Moreover, people's negative attitudes toward robots can be reduced.

Furthermore, unlike professional massage, daily home massage is characterized by nonspecialist positioning and an unspecified range of massage. As such, autonomous massage positioning faces numerous unique challenges: (1) the system should track information on key points of the human skeleton in real-time, such as the shoulders and waist, as well as hands, to enable massage tracking and (2) the system should identify massage zones and massage points based on an understanding of people's natural expressions.

Accordingly, to solve the above application pain points of intelligent robots, this study focuses on the vital issue of massage area positioning and proposes a method based on natural human-robot interaction with online learning capability.

2. Related Work

2.1. Positioning of Acupuncture Points. Existing research on massage localization has focused on the identification of body acupoints, and localization is primarily achieved using two methods, including manual-assisted finding and neural network training. The first type of manual assistance requires the marking of the target acupuncture point before the massage, mainly through the posting of the colored origin or 2D codes, after which the system locates the point by identifying the location of the marked point [5, 6]. The manually assisted methods all require a prior manual setting of the reference point and are both less intelligent. The second method, requiring the use of deep learning, is progressively becoming the focus of relevant research. Xiangping and Yudan [7] adopted a neural network model based on particle swarm optimization to train a predictive model of the relative coordinates of acupoints. Later, Sun et al. [8] located two acupuncture points on the human arm with more accuracy using a deep convolutional neural network. Chen et al. [9] used migration learning to transfer

the learned facial landmark location network to the acupoint localization network, so as to further increase the accuracy of acupoint localization. The incorporation of location accuracy metrics further increased the accuracy of positioning. The above methods have all produced satisfactory results, whereas the positioning of specific acupuncture points requires professional guidance and planning to achieve the desired results. These methods are contrast to the goal of home-based daily massage, which is primarily aimed at relaxing certain tired areas and does not exert a therapeutic effect. Thus, it is important for intelligent massage positioning systems to understand people's intentions and needs through their behavior.

2.2. Intent Understanding with Natural Interaction. Natural human-computer interaction aims to eliminate the boundaries between humans and machines to achieve smooth and natural communication between humans and computers. As research progresses, HCI tends to evolve from the initial passive interaction (e.g., command line interaction and graphical interface interaction) toward active interaction (e.g., machines actively sensing and predicting people's behavior and inferring the user's mental intent) [10, 11]. Human-computer interaction research is committed to achieving intelligent applications. To achieve this goal, a wide variety of sensors are adopted to observe people's physical behavior; human expressions, gestures, gazes [12–14], and other behaviors have also been analyzed in depth. The massage positioning of an intelligent daily massage robot is primarily dependent on the operator's intention, which can be expressed in various ways (e.g., pointing, speech, and gesture). Accordingly, the above modalities need to be considered together to analyze human intentions using contextual information [15–17]. In addition, Liu et al. [18] proposed a multitask model combining STGCN-LSTM and YOLO to recognize human intentions. Batzianoulis et al. [19] proposed the idea of determining control attribution based on people's personal preferences. Kim et al. [20] have proposed a method to identify patterns in people's daily lives that combines intention and event algorithms. Duncan K et al. [21] proposed a Markov model based on a "goal-action-intention network" through iterative Bayesian updating of the network to give it the ability to learn people's habits. Inspired by the above studies, this study adopts a recursive Bayesian algorithm to equip the system with the ability to learn.

2.3. Intent Understanding with Natural Interaction. To accurately identify finger-pointing, Smari and Salim Bouhrel [22] implemented fingertip tracking and recognition by contour detection on Kinect depth maps. Shukla et al. [23] proposed an appearance-based probabilistic target detection framework that enables the recognition of pointing gestures and the estimation of pointing directions. Barbed et al. [24] proposed a fine-grained variation of long-range pointing behavior detection using network training. However, it is not possible to obtain highly accurate, real-time finger-pointing information using the above methods. Thus, the

focus of research has gradually shifted to the recognition of key points on the body and in the hands. Although Kinect can acquire human skeletal points, it cannot be adopted for accurate pointing since it can acquire a small amount of key information. Simon et al. [25] developed a method to obtain a finger-pointing fine-grained detector through training with a multicamera system to achieve high accuracy in hand keypoint detection. Zhang et al. [26] built a multihand tracking system capable of running on the device in real time and achieving a high degree of accuracy.

In brief, existing massage systems suffer from the above key problems: (1) the system requires auxiliary conditions for manual massage positioning of massage points; (2) the system is unable to learn the massage habits of the user autonomously; and (3) the system is unable to achieve massage at any specific location. In this study, the above key scientific problems are solved by placing a focus on a multimodal intent fusion understanding approach. Combining natural pointing and speech representation, the major elements of intelligent massage systems are investigated (e.g., precise positioning and natural human-machine interaction).

3. Materials and Methods

The interaction device of the intelligent massage positioning system primarily comprises a xArm robotic arm, a Kinect perception device, a voice input device, and a computing and processing device, as illustrated in Figure 1. The difficulty of the implementation of the massage positioning system lies in how the system is capable of naturally and accurately sensing the location of the massage area expressed by people with the use of natural pointing gestures and speech. This study proposes an online learning massage positioning algorithm to solve the above difficulty. First, the idea of redundancy is used to extend the intersection of the pointing line and the body into a pointing intersection line to determine a massage candidate area. Second, an interrogative interaction is performed for the massage candidate area using roulette selection to determine the massage center point and the massage point generation model. Lastly, according to the massage area and center point, the selection probability and the central probability of the respective zone under the part are updated. After multiple selections of the same part, the system is capable of learning people's massage habits on the part, thus decreasing the number of interrogation interactions for the next center point confirmation.

To gain insights into the user's intention expressed through speech, a speech intention database, KWLib, should be first created, which describes the correlation between speech and possible intentions. The system uses real-time keyword detection for speech recognition and intent matching. In addition, to understand the user's pointing information, this system detects key points on the body and hands in real-time to achieve pointing recognition. For the intelligent massage positioning system in this study, the two modalities of speech and pointing can be either parallel inputs or single inputs. The input is assigned to three cases: the first is two modalities for parallel input, when it is

necessary to determine whether there is a contradiction between the information transmitted by both; if there is, the system will actively remind the user and ask him to re-express it. The second is when two modalities are inputted in parallel, and there is no contradiction identified between them, or only pointing serves as a single modal input. As a result, the system will turn on the OLMP algorithm to massage the localization function. Third, with voice only as the single input, the system will perform a full-area massage on the body part expressed by voice.

Pointing expressions can fall into two types, including contact and noncontact. For the first type of contact expression, the system directly employs the contact point as the center point of the massage. The understanding of the second type of noncontact expression is the difficulty and focus of this study's research. Theoretically, the intersection of the pointing line and the body can be used as the center point of the massage area. However, inaccurate detection of the skeletal points of the body and the user's own reasons (e.g., inability to raise the arm) can cause greater disturbance to the position of the intersection point. Hence, there is an error in using the intersection point as the center point of the massage. It is noteworthy that if the hand is far from the target area, a small deviation in pointing may cause the intersection point to be far from the target point. Furthermore, intelligent massage positioning requires the identification of a massage area rather than just a massage center point. The intelligent massage positioning system proposed is capable of solving the above problems of noncontact expression, and its structure is illustrated in the following diagram:

The system structure consists of three main parts (Figure 2). The first part is the area of number. 1: Through the user's basic input data to understand the intention, to determine the range of the massage candidate area, its main goal is to select a general massage area and reduce positioning errors; the second part is the area of number. 2: Through the roulette selection method to determine the interrogation point within the massage candidate area, its main goal is to determine the location of the massage center point, massage point two-dimensional (2D) distribution model, and massage The third part is the area of number. 3: Based on the user's selection results, the selection probability and central probability of the relevant body parts are constantly updated, and its main aim is to reduce the number of human-computer interactions when positioning the massage center point. The main three sections are elucidated below.

3.1. Voice Detection to Determine the Massage Part. We first build a voice intent database KWLib, which stores the set pairs of voice keywords and body part numbers. In addition to body part keywords, the system also focuses on directional words as well as negative words. Finally, all the obtained keyword information is used as the input of the intent database.

Among them, for speech recognition, we use the speech recognition technology of Baidu API to perform real-time



FIGURE 1: Interactive device diagram for intelligent massage positioning systems: (a) Kinect 2.0 devices; (b) xArm robotic arm devices.

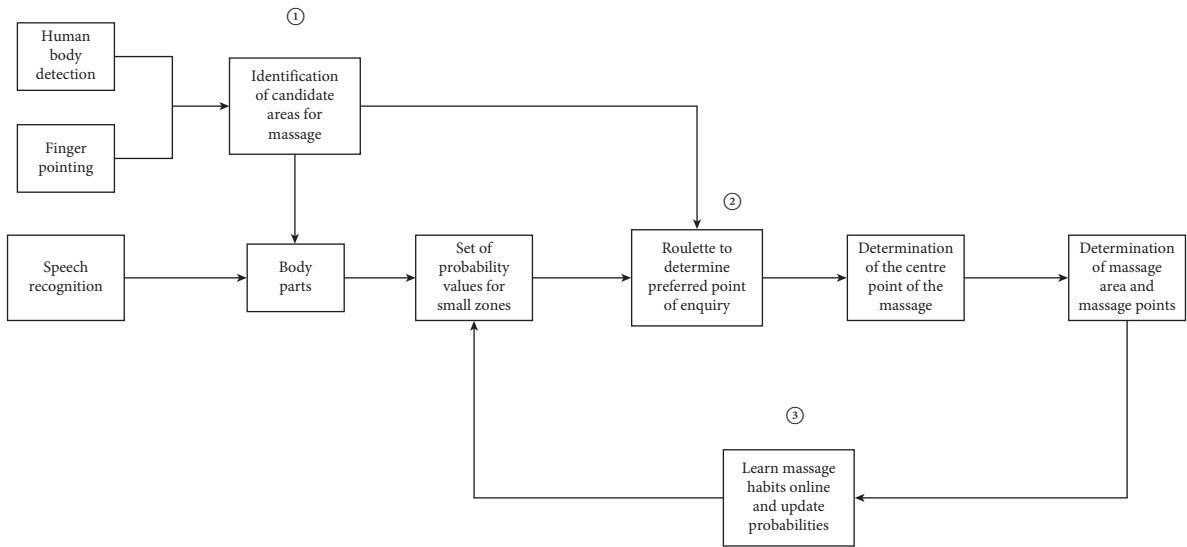


FIGURE 2: Illustration of the structure of the massage positioning system.

speech detection. After the system performs the massage positioning scenario, the system detects the user's voice expression in real time. When the body part keyword is detected, the final body part number is obtained by combining the before and after information. The identification process is shown in Figure 3.

3.2. Natural Pointing to Identify Candidate Area for Massage. After considering accuracy, stability, and real time, this study uses the research results of Zhang F et al. and Bazarevsky et al. [27] on hand and body key points as the method of acquiring the base data. The underlying data are processed to identify the candidate area.

First, it is considered that there may be irregularities in user pointing gestures. To increase accuracy, pointing is assigned to two cases. The first case is when the index finger is bent during the pointing process; the second case is when the index finger is not bent during the pointing process. The system sets different pointing lines in accordance with the different cases.

Second, once the pointing line, $Line_1$, has been determined, we assume that there exists a surface α . The straight

line, $Line_1$, lies within α and that α is perpendicular to the ground (the xoz face in 3D space). This study translates the above explicit conditions into mathematical form: the normal \mathbf{n} of the ground is known to be $(0 \ 1 \ 0)$ and let the direction vector \mathbf{l}_1 of the line, $Line_1$, be $(a \ b \ c)$. With geometric knowledge, the normal vector \mathbf{n}_α of the surface α is expressed as follows:

$$\mathbf{n}_\alpha = \mathbf{n} \times \mathbf{l}_1 = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 1 & 0 \\ a & b & c \end{vmatrix}. \quad (1)$$

The equation of surface α is obtained by combining the normal vector \mathbf{n}_α and the coordinates of the fingertip point. Afterward, the body plane α_{body} is obtained from the information on the coordinates of the key points pointing to the body part where the intersection point $p_{intersection}$ is located.

Lastly, the length and width of surface α and surface α_{body} are defined by the position of the pointing hand, the direction of pointing, and the body posture. Afterward, the intersection line I between the two surfaces can be found

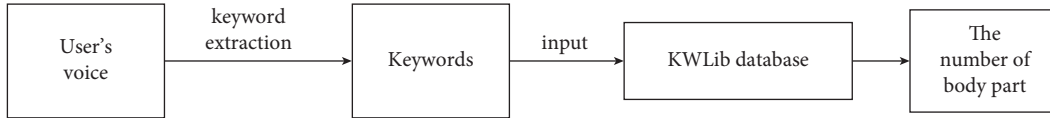


FIGURE 3: Schematic diagram of the speech recognition process.

setting the height of the massage candidate area T to the length H of the projection of I on the y -axis.

When the user's pointing action is not standard, the system uses the second finger node and the tip of the index finger as the key points of the pointing line, as shown in Figure 4(a). However, when the user's pointing action is standard, the system uses the heel and tip of the index finger as the key points of the pointing line, as presented in Figure 4(b). The red dotted line in the diagram represents the intersection line I of the two faces.

The width of the area T is obtained by d that changes with the distance d_j between the fingertip and $p_{\text{intersection}}$. If the fingertip is farther away from $p_{\text{intersection}}$, the intersection point determined by pointing may deviate significantly from the actual target point. Thus, the width of the candidate area T should be widened to maximally include the target point within it. Through extensive experimental testing, this study sets the value of d for three cases: first is $2\text{ cm} < d_j < 8\text{ cm}$; that is, the distance between the fingertip and $p_{\text{intersection}}$ is relatively close, and d is set to one quarter of the maximum value L_1 of the width of the part where $p_{\text{intersection}}$ is located; second is $8\text{ cm} < d_j < 20\text{ cm}$, where d is set to $(1/3)L_1$; third is $d_j > 20\text{ cm}$; that is, the distance between the fingertip and $p_{\text{intersection}}$ is relatively far, and d is set to $(2/3)L_1$. Afterward, the length of the projection of d and I on the x -axis is compared, and the maximum value is selected as the width L of T . The area framed by the dashed line in Figure 5(a) is the candidate area T , and the black line within this area is the intersection line I .

3.3. Determination of Massage Center Point and Massage Point Generation Model in Candidate Area. The results and discussion may be presented separately, or in one combined section, and may optionally be divided into headed subsections.

Before the system was run, this thesis first divided the body into parts, such as the left and right arms, and the back. Second, the respective part was then divided into more refined zones. The respective small zone has a selection probability value and a central probability value, and both two probabilities of the respective small zone under the same part sum to 1. The following is an example of area T falling on the back, assuming that the back contains a total of 9 small zones. Then, the determination of the massage point centroid and the massage point generation model is shown below.

Suppose the area T contains a total of m sections with area values of S_1, S_2, \dots, S_m , thus forming the set of areas \mathbf{S} . The selection and central probabilities are obtained for the respective part, resulting in a probability set θ and a probability set μ , as shown in Figure 5(b). Taking into account the

existence of user pointing bias and to avoid the smaller combined probability sections being simply ignored, this study used the roulette selection method to determine the preferred interrogation points within area T .

First, the system determines the combined probability for the respective section in area T according to (2) and calculates the cumulative probability value $Q(P_i)$ for the respective section in order. The cumulative probability for the respective section is the sum of its own probability and the probabilities of all sections that lie before it. The cumulative probability uses line segments of different lengths to represent the probability of the respective section. All sections are integrated to form a long line of length 1.

$$P(\text{PT}_i) = \frac{S_i \times \theta_i \times \mu_i}{\sum_{j=0}^m S_j \times \theta_j \times \mu_j}. \quad (2)$$

Next, the system generates a random number in the interval $[0, 1]$. The number is judged to fall within which line segment, so the preferred section of the area T is determined. Notably, the probability of a random number falling in a longer line segment is relatively high. However, there is also the possibility of shorter line segments being selected. Thus, the phenomenon of a fixed range of massage center points is avoided.

The above steps lead to the preferred interrogation section and the position of its center (x', y') within the area T . Afterward, the massage arm moves to this center point and asks the user "whether the point currently touched is included in the massage area." If the system gets a negative answer, the preferred interrogation part will be removed from the candidate area T , and the remaining sections will be used as a new candidate area T' . The system will then recalculate the combined probability value for the respective section of the area T' and use the above steps to reselect the next section. If a positive answer is obtained from the user, the point (x', y') is moved on the x -axis to the intersection line I to get a new point (x_0, y_0) , and the point serves as the massage center point, as presented in Figure 5(c). In addition, the system sets a minimum area value β for the candidate area. When the candidate area is being narrowed down, if the area of the T' is smaller than β , the system will ask the user to re-express it.

Lastly, the position coordinates of the massage points are set to be consistent with a normal distribution on the X and Y axes, and the parameters x and y are independent of each other. x_0 and y_0 are the means of the two normal distributions, respectively, and the variances are determined by L and H , respectively. In accordance with the 3σ principle, the variance of X and Y can be found as $\sigma_x = (L/6)$ and $\sigma_y = (H/6)$, respectively. Accordingly, the equation for the

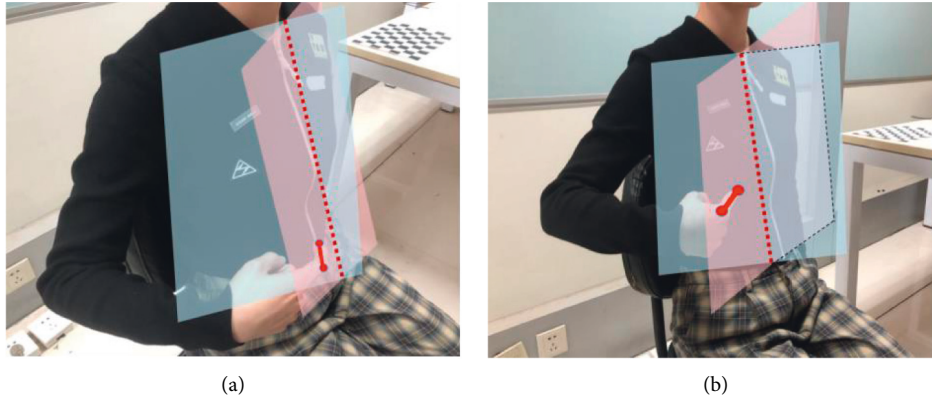


FIGURE 4: Intersection line finding diagram with different pointing lines: (a) irregular pointing; (b) regular pointing.

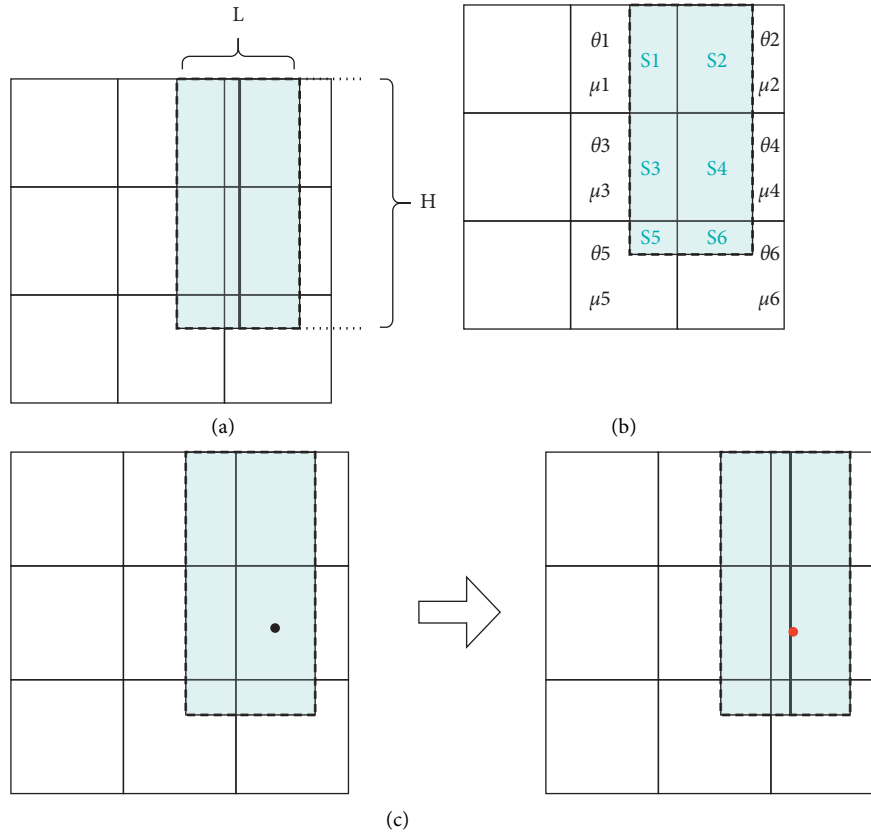


FIGURE 5: Schematic diagram of the message center point determination process: (a) the area framed by the dotted line represents T ; (b) the area, central probability, and selection probability corresponding to each small section contained in the area T is indicated; (c) the message center point is found by moving the interrogation point.

2D normal distribution that the message point coordinates obey is expressed as follows:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x-x_0^2}{\sigma_x^2} - \frac{y-y_0^2}{\sigma_y^2}\right). \quad (3)$$

The system is capable of generating the coordinates of the message points randomly according to equation (3). In addition, the area enclosed by the four points $x_0 - 2\sigma_x$,

$x_0 + 2\sigma_x$, $y_0 - 2\sigma_y$, and $y_0 + 2\sigma_y$ is set as the target area T_t for this message positioning, as shown in the area framed by the dashed line in Figure 6.

3.4. Updating the Central and Selection Probabilities Using Recursive Bayes. To make the results of intention understanding more accurate, the system should be able to learn continuously. In the case of message positioning,

intelligence means that after a number of positions, the system should learn the operator's preferences to achieve more accurate and rapid positioning. To achieve this, this study requires that the system learn the user's historical massage area and massage frequency and be able to automatically update the selection probability values and the central probability values for each small zone under the relevant part. The above goal can be achieved using a recursive Bayesian approach. Suppose that the R th part of the body is selected N times and that the part covers a total of K small zones. The result x_R^i of the i th selected massage positioning is expressed as $x_R^i = (Q_{R1} = 0, Q_{R2} = 0, \dots, Q_{Ri} = 1, \dots, Q_{RK} = 0)$, with Q_{Ri} representing the i th small zone under the R th part. A value of 0 for Q_{Ri} means that the central point of massage does not fall in the i th small zone; a value of 1 for Q_{Ri} means that the central point of massage falls in the i th small zone. In addition, each of the K small zones has a central probability value and a selection probability value, which can form the probability set $\theta_R = \{P_{R1\text{-center}}, P_{R2\text{-center}}, \dots, P_{RK\text{-center}}\}$ and $\mu_R = \{P_{R1\text{-selection}}, \dots, P_{RK\text{-selection}}\}$.

Online learning aims to update the central and selection probabilities of the respective small zone under the relevant part using the results of the user's massage positioning selection, that is, to update the probability set θ_R and μ_R .

Using Bayes' formula, the posterior probability of the central probability can be written as follows:

$$P(\theta_R | x^i) \propto P(x^i | \theta_R) \cdot P(\theta_R | x^{i-1}). \quad (4)$$

According to (4), the system is capable of understanding user preferences based on continuous learning. The prior function $P(\theta_R | x^{i-1})$ can be obtained by iterating step by step through $P(\theta_R | x^{i-2}), \dots, P(\theta_R | x^0)$, that is, $P(\theta_R)$. $P(\theta_R)$ is the initial probability distribution for the small zone. We set it to obey the Dirichlet distribution, and hence, the posterior probabilities also obey the Dirichlet distribution. The likelihood function $P(x^i | \theta_R)$ can be obtained by the following equation:

$$P(x^i | \theta_R) = \prod_{j=1}^K P(x_j^i | \theta_R) \propto \prod_{j=1}^K \theta_{Rj}^{Q_{Rj}}. \quad (5)$$

Hence, the maximum a posterior estimate of θ_R is given by the statistically large amount of positioning data, as shown in the following equation:

$$\widehat{\theta}_{Rj} = \frac{Q_{Rj} + \alpha_{Rj} - 1}{\sum_{i=1}^K Q_{Ri} + \sum_{i=1}^K (\alpha_{Rj} - 1)}, \quad (6)$$

where α_{Rj} is a Dirichlet parameter that records the prior counts of the observed massage centroids falling in the j th zone.

Finally, the selection probabilities are progressively updated according to the way the central probabilities are updated, as shown in the following equation.

$$\widehat{\mu}_{Rj} = \frac{C_{Rj} + HS_{Rj} + s_{Rj}}{\sum_{i=1}^K C_{Ri} + \sum_{i=1}^K HS_{Ri} + \sum_{i=1}^K s_{Ri}}, \quad (7)$$

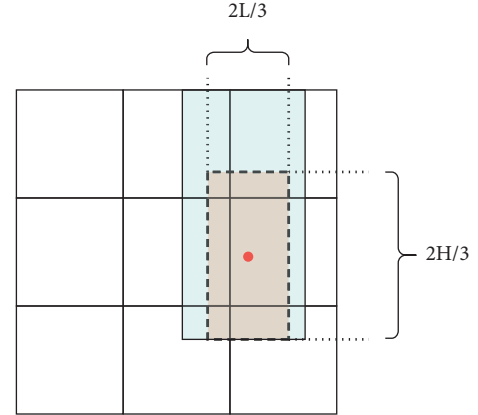


FIGURE 6: Schematic diagram of the massage area determination.

where HS_{Rj} denotes the sum of the areas of the j th small zone contained within the target area of the historical massage; C_{Rj} represents the area of the j th small zone contained within the current target massage area; and s_{Rj} is the area of the j th small zone. Thus, the selection probability value μ_{Rj} for the respective small zone on the R th part will be progressively updated as an increasing amount of interaction information is added.

3.5. Online Learning Massage Positioning Algorithm.

Based on the above discussion, the basic idea of the online learning massage positioning algorithm (OLMP) is elucidated below. (1) When the system detects the noncontact pointing gesture, it exploits the redundancy idea to extend the pointing intersection as the intersection line to determine the massage candidate area T . (2) The system determines the massage center point using the roulette selection method and then determines the 2D normal distribution model of the massage point in accordance with the height and width of the area T . (3) In accordance with the massage target area, the selection and central probabilities of relevant body parts are updated to progressively realize the function of learning the user's massage habit. The algorithm in this study is defined as Algorithm 1.

3.6. Algorithm Analysis.

The main features exhibited by the OLMP algorithm are elucidated below. (1) The OLMP algorithm can understand the user's representation of any body's position under natural pointing. The system determines the user's pointing direction in accordance with the key point of the finger. If the pointing line does not intersect with the body (e.g., the pointing line points to the outside of the body), the system will actively remind the user and ask him/her to re-express it. If there is an intersection between the pointing line and the body, the system will be consistent with the steps in Section 3.1 to find the candidate area T and determine the center point of the massage within the area T according to Section 3.2. During the above process, the system records the position of the massage point in relation to the body's key points. All the above settings ensure that the massage point is always found when the pointing

Input: The user's voice: voice; the key points of fingers; the key points of body; the collections of two types of probability for the respective part of the body.

Output: The collection \mathbf{P} of message points.

- (1) Computer \mathbf{n}_α using the formula (1);
 $\alpha \leftarrow \mathbf{n}_\alpha$ and Line_1 ; $\alpha_{\text{body}} \leftarrow P_{\text{intersection}}$ and the key points of body;
 $I \leftarrow \alpha$ and α_{body} ./* calculate the intersection of the surface α and α_{body} .*/
- (2) $H = \text{height}(I)$; $L = \max(\text{Length}(I), d)$.
- (3) $T \leftarrow (H, L, I)$ /* determine the message candidate area T .*/
- (4) $\text{Num1} \leftarrow \text{KWLlib}(\text{voice})$; /* match the user's voice to the speech intent library to find the number of the target body part */
 $\text{Num2} \leftarrow T$./* determine the number of the body part where the candidate area T is located. */
- (5) IF $\text{Num1} = \emptyset$ and $\text{Num2} \neq \emptyset$
 $\text{num} = \text{Num2}$
 IF $\text{Num1} \neq \emptyset$ and $\text{Num1} \cap \text{Num2} \neq \emptyset$
 $\text{num} = \text{Num1} \cap \text{Num2}$.
 /* determine the number of the target part. */
- (6) The sets of initial probabilities for the respective part within T :
 $\theta \leftarrow \theta_{\text{num}} \cap T$; $\mu \leftarrow \mu_{\text{num}} \cap T$.
- (7) Compute $F(\text{PT}_i)$ using formula (2).
- (8) Center point $(x_0, y_0) \leftarrow \text{PT}_{\text{chosed}}$ /* roulette selection ($F(\text{PT}_i)$)./* finding the section $\text{PT}_{\text{chosed}}$ through the roulette selection method. Determining the center point (x_0, y_0) of target message area by $\text{PT}_{\text{chosed}}$. */
- (9) Compute $f(x, y)$ using the formula (3).
- (10) $\mathbf{P}, \mathbf{T}_t \leftarrow f(x, y)$./* determine the set of message points \mathbf{P} and the target area for message \mathbf{T}_t . */
- (11) Update θ_{num} using the formula (6).
- (12) Update μ_{num} using the formula (7).

Output \mathbf{P}
 END

ALGORITHM 1: Online learning message positioning (OLMP).

information is correct. (2) The selection and central probabilities of the respective zone under the body part can be constantly updated online to learn the user's massage habits.

The main differences between the OLMP algorithm and existing methods are elucidated below. (1) The algorithm is capable of achieving message localization without the need for other auxiliary conditions by analyzing the user's non-contact pointing expression of the message area under natural conditions. The above function reduces the user's memory and operational load. (2) The algorithm can update the probability value of the respective small zone under the relevant part based on the target area obtained from the localization. This function allows the system to find the message center point from the message candidate area T more rapidly in the next positioning, reducing the number of times the system asks the user. (3) In the positioning process, the user can move his body as he pleases without being restricted to a single posture. The above system will achieve real-time tracking of the message points based on the recorded location of the message area points in relation to the key points of the body.

4. Experimental Results and Analysis

The proposed OLMP algorithm is integrated with the xArm robotic arm for intelligent message positioning. In this section, the effectiveness and reliability of this proto-type system are further verified, and the intention understanding rate and cognitive load of the algorithm are evaluated.

4.1. Experimental Settings. The intelligent positioning system in this study comprises a Kinect device, a computer with an I7-10875H CPU, an RTX2060 GPU, 16G of RAM, and the xArm 7-axis robotic arm. To be specific, the arm was fixed to the table, and the Kinect camera was fixed to the right of the arm, which was approximately 1.2 m away from the user. 20 volunteers between the ages of 35 and 65 were invited to the experiment at a male to female ratio of 1 : 1.

Since the massage arm has a limited range of movement, to achieve effective message positioning, the area in which the user intends to express himself should be limited, and the user should choose the posture in which the massage can be performed. Moreover, due to the difficulty of expressing the back area by pointing, a special condition was set; that is, the experimenter could point to the front chest, instead of pointing to the back area, and the system would automatically map the front area to the back. However, the experimenter should first verbalize the part of the body that he or she wants to massage. Indeed, the experimenter can also express his or her intention directly to an area of the back.

Furthermore, the types of user responses are classified into positive and negative responses. The keywords of positive responses consist of "yes," "right," "correct," and others, while the keywords of negative responses comprise "no," "not in," "negative," "none," etc.

4.2. Experimental Procedure. The respective experimenter should perform 20 repetitive message area selections with natural pointing. When the experimenter enters the

designated area to express their intention using pointing, if there is a problem with their pointing (e.g., pointing in a direction unrelated to their body), or if there is a contradiction in the parallel input of speech and pointing, the experimenter is asked to repeat the expression, and no count is made in either case. The experimenter can point in a variety of postures (e.g., standing and sitting), as illustrated in Figure 7. If the experimenter is in an area that is out of reach of the robotic arm, the system will alert the user to make position adjustments.

Figure 8 illustrates the whole process of massage localization. Once the pointing gesture is detected and the body posture is stable, the system will determine the candidate area T , which has a height of 200 pixels and a width of 110 pixels, as shown in Figure 8(a). To prevent the body from moving, the system records the position relationship between the massage candidate area points and the rest of the body's key points. Afterward, the OLMP algorithm is used to detect which parts of the area T contain small zones and to determine the preferred interrogation point, as shown in Figure 8(b). The robotic arm moves to this point and initiates the interrogation: does this contact location fall within the target area, as shown in Figure 8(c). When an affirmative answer from the user is detected, the point is moved to the intersection line to obtain the massage center point, and a 2D distribution model of the massage point coordinates is obtained according to equation (3) as follows:

$$(x, y) = \frac{9}{11000\pi} \exp\left(-\frac{9(x-574)^2}{3025} - \frac{9(y-394)^2}{10000}\right). \quad (8)$$

The system determines the massage target area and the set of massage points based on the 2D distribution model above, as shown in Figure 8(d), and records the location of the massage points in relation to key points on the body. Ultimately, the probability values for the selection and the central of the respective small zone under the relevant part are updated during the positioning process, the experimenter can move or change posture, and the system can achieve real-time tracking of the massage.

4.3. Experimental Results. In this study, the proposed OLMP algorithm was validated and evaluated in terms of three metrics, including accuracy, number of interactions, and user cognitive load.

4.4. Accuracy. The accuracy of massage positioning can be derived by the following equation:

$$\text{Accuracy Rate} = \frac{\text{Count}}{20}, \quad (9)$$

where Count denotes the number of correct massage points in the set \mathbf{p} of massage points. The counting method is that the robot arm moves to the position of the massage point in \mathbf{p} , respectively, and asks: "Is the current contact point located in the area you want to massage." If the answer is positive, the count of Count increases by 1. Otherwise, the count of does not increase.

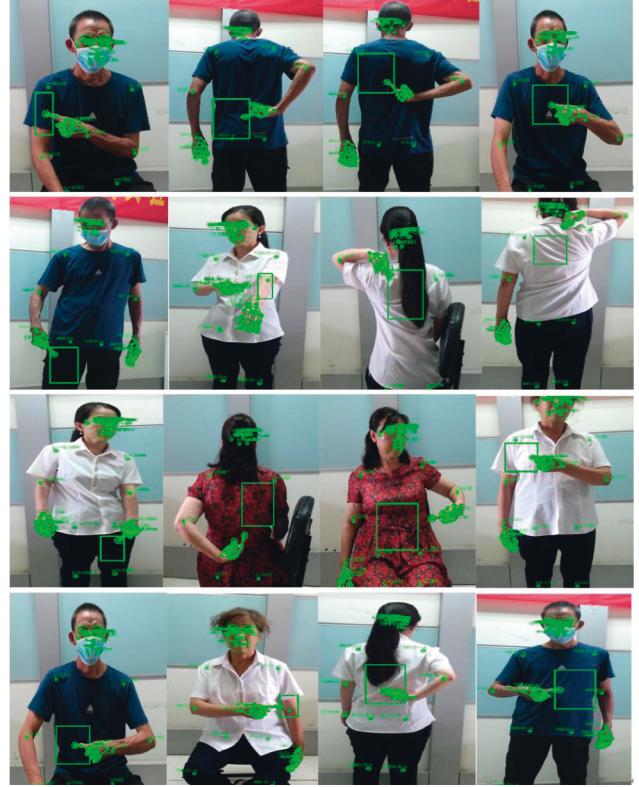


FIGURE 7: Collection of massage positioning charts.

The results were recorded for 20 experimenters' 20 localization massages, that is, 400 times of intention understanding. The accuracy rate under 400 selections was counted, and the results are presented in Figure 9.

As depicted in the figure, 291 out of 400 intention comprehension tasks achieved a correct rate of 70% or more, for a total of 72.5%. In addition, the cases where the correct rate was below were counted and analyzed, and it was found that the case mainly occurred when the user points directly at the back area with their finger. Due to the limitations of people's limb range of movement, the experimenter's pointing direction and the target area can deviate significantly. The above phenomenon decreases the correct rate of intention understanding.

4.5. Number of Interactions. The main innovation of the OLMP algorithm is that the system is capable of learning the user's massage preferences online and decreasing the number of queries for the next massage positioning.

To verify the effectiveness of this innovation, two volunteers were randomly selected and asked to make ten repeatable choices for their back and waist distributions, and the choices should meet their massage needs. To avoid interfering between choices, the experimenter was asked to rest for 10 min after the respective selection was made. When the experimenter is changed, the system resets the initial probability values for the respective area to learn the user's massage habits more rapidly. Furthermore, the system is set to contain 16 small zones on the back of the body and 9 small zones on the waist.

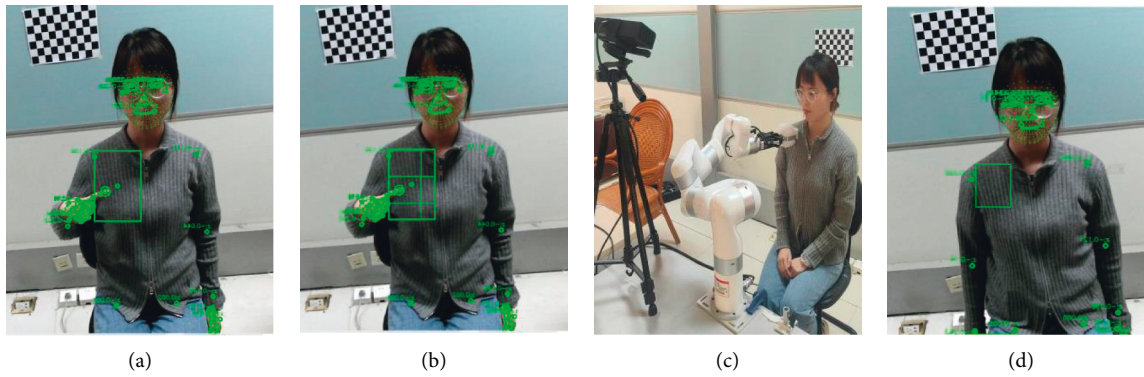


FIGURE 8: The massage area determination process: (a) the candidate area T is determined by pointing lines; (b) the area, selection probability, and central probability of the small area contained in area T are found; (c) the robot arm moves to the preferred point determined by Equation (2) and initiates a query to the user; (d) a positive response is obtained from the user, which leads to the determination of the massage area and the massage point.

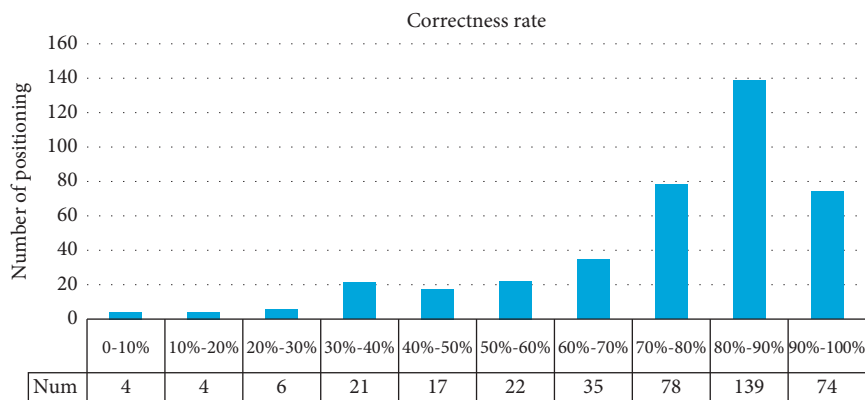


FIGURE 9: Distribution of correct rates.

The number of interactions in Figure 10(a)–10(c) implies the number of times the system controlled the robot arm to move to the target point to initiate a query to the user when the center point of the massage within T is being determined. As revealed by the graphs, both experimenters performed a relatively high number of interrogation interactions during the first positioning process. With the increase in the number of times the experimenter positioned the same part, the number of interrogations required decreased. The average number of interactions (c) suggested that after six selections, only one interrogation interaction was generally required to locate the center of the massage. For the random selection model, the number of interrogation interactions does not decrease with the increase in the number of experimenter orientations. Thus, the OLMP algorithm can decrease the number of human-machine interactions during the positioning process by learning the user's massage habits online.

4.6. NASA-TLX User Reviews. All 20 experimenters were invited to complete a NASA Task Load Index questionnaire after the experiment was performed. The questionnaire consisted of six evaluation indicators below, including time

demand (TD): the efficiency with which time is managed during the experiment; physical demand (PD): the level of physical effort demanded by the experiment; personal performance (OP): the level of self-satisfaction in completing the experiment; energy (E): the amount of effort required to achieve the self-assessed level; and frustration (F): how you feel throughout the experiment.

The NASA-TLX generally comprises two steps. Step 1: a two-by-two comparison of the six indicators in 15 sets. The experimenters were allowed to weigh in and select one indicator at a time to calculate the relevance of the indicators to the task. The results are illustrated in Figure 6 as data widths. Step 2: the respective indicator was scored, where the respective indicator was divided into 5 equal intervals, the respective in increments of 1, with 5 as the maximum. The mean values of the correlations between the factors in the 20 questionnaires were derived, as well as the mean value of the respective factor score. The mean variance was obtained as 0.9 for the mental factor, 0.592 for the physical factor, 0.943 for the time factor, 0.843 for the satisfaction factor, 1.122 for the energy factor, and 0.81 for the frustration factor, as illustrated in Figure 11.

As revealed by the statistical results, the OP factor has the highest effect. On the basis of the above factor, the OLMP

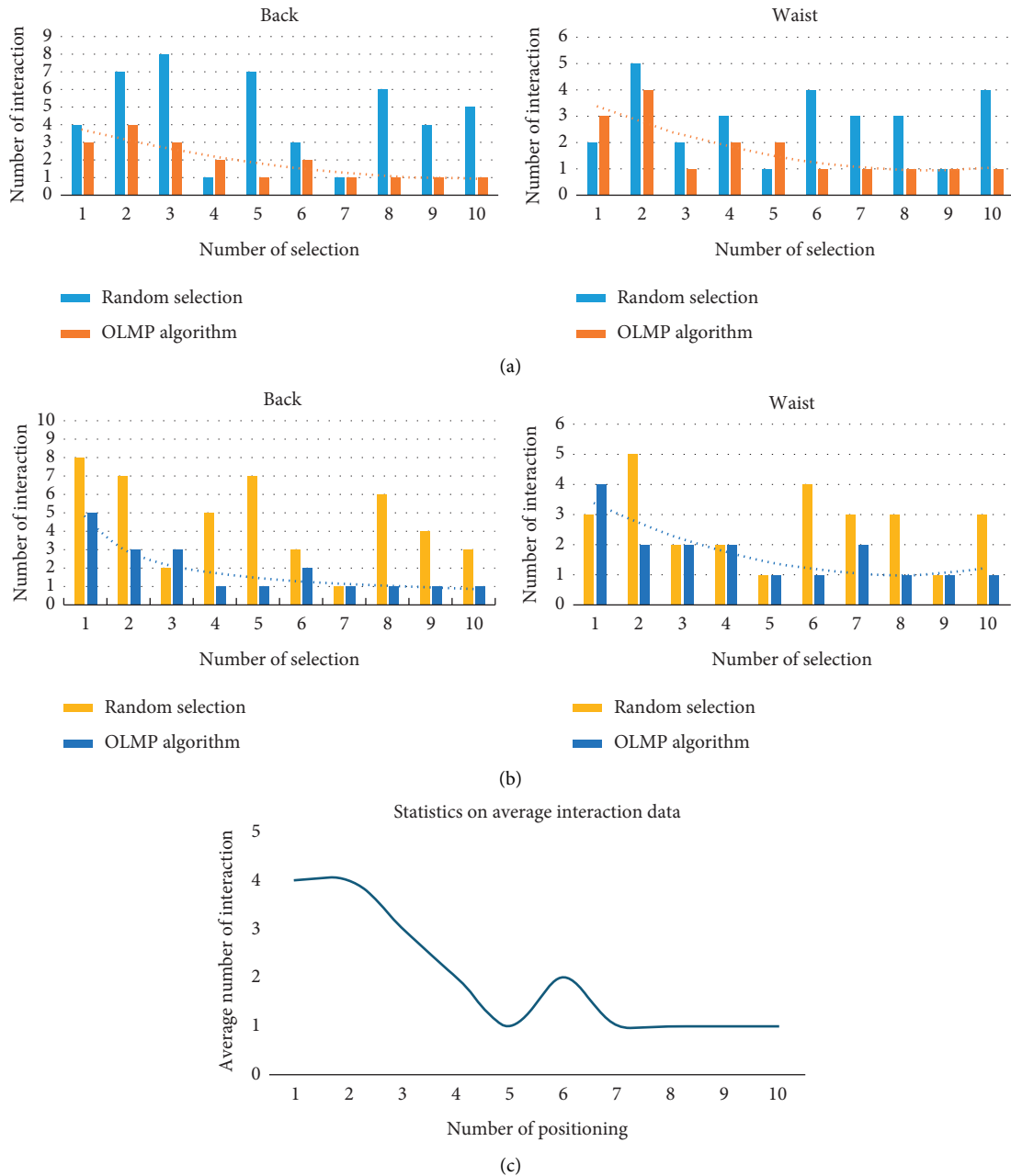


FIGURE 10: Collection of experimental results plots for the online learning function: (a) statistical plots of the number of interactions for #1 experimenter for 10 back and lumbar massage orientations, respectively; (b) statistical plots of the number of interactions for #2 experimenter for 10 back and lumbar massage orientations, respectively; (c) change of the number of interrogative interactions required during online learning.

algorithm is slightly better than the positioning mode of the massager. This finding can be explained below. Since massage chairs have a relatively fixed massage area, they are not sufficiently flexible to target a small area in accordance with the user’s wishes. However, the positioning of the massage under natural pointing can be more consistent with the psychological needs of the user, so the OLMP algorithm can bring a higher level of satisfaction to the user. Moreover, as depicted in the graph, the natural pointing and voice expressions carry less load than the operating panel or the

remote control. Lastly, a weighted calculation of the six factors gives a total load value of 1.92 for the OLMP algorithm positioning and 2.00 for the massage chair positioning. In brief, the OLMP algorithm positioning has better application prospects than the operator board positioning and also illustrates the effectiveness of the OLMP algorithm. Furthermore, in this study, the age-segmented statistics of the questionnaire was analyzed, and it was found that the OLMP algorithm positioning was much more favorably received by the elderly than the massage chair positioning in

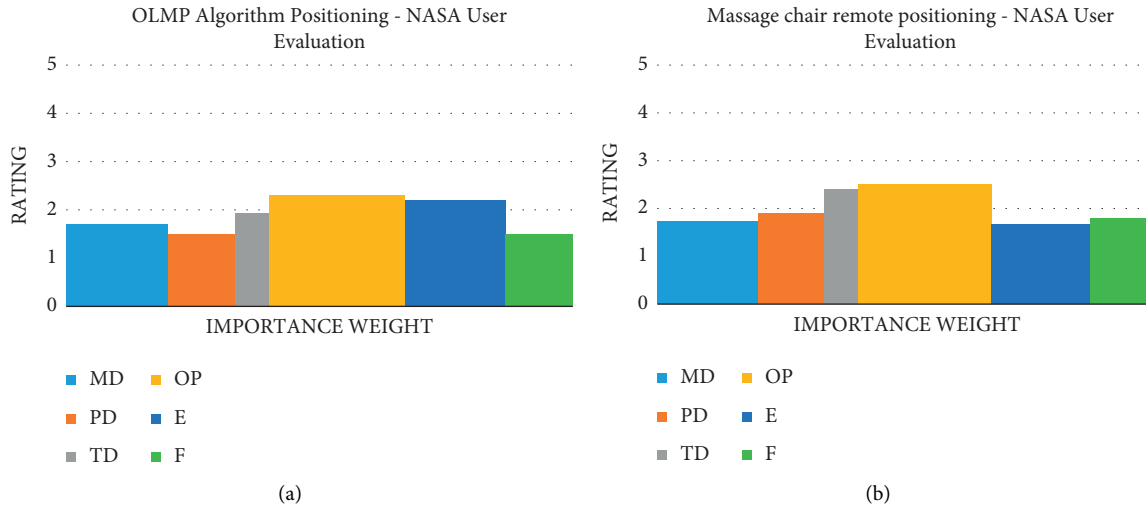


FIGURE 11: NASA-TLX user evaluation graphs: (a) statistical graph of user evaluations for OLMP algorithm positioning; (b) statistical graph of user evaluations for massage chair positioning.

terms of the mental factor and the self-satisfaction factor, while the young people's evaluation of the two methods was mixed. As revealed by the above analysis, the OLMP algorithm's core concept can be integrated into elderly assistance and escort robots for intelligent applications.

5. Conclusions

In this study, existing massage positioning methods are analyzed, and the OLMP algorithm is proposed by combining the concepts of natural interaction and precise positioning. The OLMP algorithm is to essentially integrate iterative Bayesian online learning of people's daily massage habits for accurate positioning of the massage area with a small amount of interaction. As revealed by the results of the experiments, massage positioning based on the OLMP algorithm can be achieved naturally and in real time without the need for any auxiliary tools and can reduce the memory and operational load on people.

Data Availability

The source code data used to support the findings of this study have not been made available because the source code belongs to laboratory assets, and we cannot open source without authorization.

Conflicts of Interest

The authors declared that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This paper was supported by the Independent Innovation Team Project of Jinan City (No. 2019GXRC013) and the Shandong Provincial Natural Science Foundation (No. ZR2020LZH004).

References

- [1] C. Granata, A. Ibanez, and P. Bidaud, "Human activity-understanding: a multilayer approach combining body movements and contextual descriptors analysis," *International Journal of Advanced Robotic Systems*, vol. 12, no. 7, p. 89, 2015.
- [2] M. W. Alibali, "Gesture in spatial cognition: expressing, communicating, and thinking about spatial information," *Spatial Cognition and Computation*, vol. 5, no. 4, pp. 307–331, 2005.
- [3] G. Butterworth and S. Itakura, "How the eyes, head and hand serve definite reference," *British Journal of Developmental Psychology*, vol. 18, no. 1, pp. 25–50, 2000.
- [4] S. Mayer, V. Schwind, R. Schweigert, and N. Henzex, "The effect of offset correction and cursor on mid-air pointing in real and virtual environments," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, New York, NY, USA, April 2018.
- [5] K. Sun, Q. Zhao, Z. Yang, and X. Xu, "Visual Feedback System for Traditional Chinese Medical Massage Robot," in *Proceedings of the 2019 Chinese Control Conference (CCC)*, pp. 6379–6385, Guangzhou, China, July 2019.
- [6] S. Lin and P. Yi, "Human acupoint positioning system based on binocular vision," *IOP Conference Series: Materials Science and Engineering*, vol. 569, no. 4, Article ID 042029, 2019.
- [7] Y. Xiangping and W. Yudan, "Acupoint positioning system based on pso-bp neural network," *Application of Electronic Technique*, vol. 09, 2018.
- [8] L. Sun, S. Sun, Y. Fu, and X. Zhao, "Acupoint detection based on deep convolutional neural network," in *Proceedings of the 2020 39th Chinese Control Conference (CCC)*, pp. 7418–7422, Shenyang, China, July 2020.
- [9] Y. Chen, H. Yang, D. Chen, and X. Chen, "Facial Acupoints Location Using Transfer Learning on Deep Residual Network," in *Proceedings of the 2021 7th International Conference On Computer And Communications (ICCC)*, pp. 1842–1847, Chengdu, China, December 2021.
- [10] E. B. Sandoval, J. Brandstetter, M. Obaid, and C. Bartneck, "Reciprocity in human-robot interaction: a quantitative approach through the prisoner's dilemma and the ultimatum game," *International Journal of Social Robotics*, vol. 8, no. 2, pp. 303–317, 2016.

- [11] T. Chaminade, D. Rosset, D. Da Fonseca et al., "How do we think machines think? An fMRI study of alleged competition with an artificial intelligence," *Frontiers in Human Neuroscience*, vol. 6, p. 103, 2012.
- [12] M. R. Fraune, S. Sherrin, S. Šabanović, and E. R. Smith, "Is Human-robot interaction more competitive between groups than between individuals," in *Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 104–113, Daegu, Korea(South), March 2019.
- [13] M. Belkaid, K. Kompatsiari, D. De Tommaso, I. Zabliith, and A. Wykowska, "Mutual gaze with a robot affects human neural activity and delays decision-making processes," *Science Robotics*, vol. 6, no. 58, Article ID eabc5044, 2021.
- [14] Z. Hu, Y. Zhang, Y. Xing, Y. Zhao, D. Cao, and C. Lv, "Toward human-centered automated driving: a novel spatial-temporal vision transformer-enabled head tracker," *IEEE Vehicular Technology Magazine*, pp. 2–9, 2022.
- [15] A. Graser, T. Heyer, L. Fotoohi et al., "A supportive friend at work: robotic workplace assistance for the disabled," *IEEE Robotics and Automation Magazine*, vol. 20, no. 4, pp. 148–159, 2013.
- [16] M. Shishehgar, D. Kerr, and J. Blake, "The effectiveness of various robotic technologies in assisting older adults," *Health Informatics Journal*, vol. 25, no. 3, pp. 892–918, 2019.
- [17] Z. Hu, Y. Xing, W. Gu, D. Cao, and C. Lv, "Driver anomaly quantification for intelligent vehicles: a contrastive learning approach with representation clustering," *IEEE Transactions on Intelligent Vehicles*, p. 1, 2022.
- [18] C. Liu, X. Li, Q. Li, Y. Xue, H. Liu, and Y. Gao, "Robot recognizing humans intention and interacting with humans based on a multi-task model combining ST-GCN-LSTM model and YOLO model," *Neurocomputing*, vol. 430, no. 3, pp. 174–184, 2021.
- [19] I. Batzianoulis, F. Iwane, S. Wei et al., "Customizing skills for assistive robotic manipulators, an inverse reinforcement learning approach with error-related potentials," *Communications biology*, vol. 4, no. 1, p. 1406, 2021.
- [20] J. M. Kim, M. J. Jeon, H. K. Park, S. H. Bae, S. H. Bang, and Y. T. Park, "An approach for recognition of human's daily living patterns using intention ontology and event calculus," *Expert Systems with Applications*, vol. 132, pp. 256–270, 2019.
- [21] K. Duncan, "Scene-dependent human intention recognition for an assistive robotic system," in *Proceedings of the European Conference on Computer Vision*, Springer, Berlin, Germany, September 2014.
- [22] K. Smari and M. Salim Bouhlel, "Gesture recognition system and finger tracking with kinect: steps," in *Proceedings of the 2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, pp. 544–548, Hammamet, Tunisia, December 2016.
- [23] D. Shukla, O. Erkent, and J. Piater, "Probabilistic detection of pointing directions for human-robot interaction," in *Proceedings of the 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–8, Adelaide, Australia, November 2015.
- [24] O. L. Barbed, P. Azagra, L. Teixeira, and M. Chli, J. Civera, A. J. Murillo, "Fine-grained pointing recognition for natural drone guidance," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1040–1041, Seattle, WA, USA, June 2020.
- [25] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand keypoint detection in single images using multiview bootstrapping," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1145–1153, Honolulu, HI, USA, July 2017.
- [26] F. Zhang, V. Bazarevsky, A. Vakunov et al., "Mediapipe hands: on-device real-time hand tracking," 2020, <https://arxiv.org/abs/2006.10214>.
- [27] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "Blazepose: on-device real-time body pose tracking," 2020, <https://arxiv.org/abs/2006.10204>.

Research Article

A Novel Fault Diagnosis Method for Denoising Autoencoder Assisted by Digital Twin

Wenan Cai ¹, Qianqian Zhang ² and Jie Cui³

¹School of Mechanical Engineering, Jinzhong University, Jinzhong 030619, China

²School of Automation and Software Engineering, Shanxi University, Taiyuan 030006, China

³School of Mechanical Engineering, North University of China, Taiyuan, Shanxi 030051, China

Correspondence should be addressed to Wenan Cai; caiwenan65@163.com

Received 14 May 2022; Accepted 5 July 2022; Published 21 July 2022

Academic Editor: Jie Liu

Copyright © 2022 Wenan Cai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Digital twin (DT) is an important method to realize intelligent manufacturing. Traditional data-based fault diagnosis methods such as fractional-order fault feature extraction methods require sufficient data to train a diagnosis model, which is unrealistic in a dynamically changing production process. The ultrahigh-fidelity DT model can generate fault state data similar to the actual system, providing a new paradigm for fault diagnosis. This paper proposes a novel digital twin-assisted fault diagnosis method for denoising autoencoder. First, in order to solve the problem of limited or unavailable fault state data for machines in dynamically variable production scenarios, a DT model of the machine is established. The model can simulate a dynamically changing production process, thereby generating data for different failure states. Second, a novel denoising autoencoder (NDAE) with Mish as the activation function is proposed and trained using the source domain data generated by DT. Finally, in order to verify the effectiveness and feasibility of the proposed method, the method is applied to a fault diagnosis example of a triplex pump, and the results show that the method can realize intelligent fault diagnosis when the fault state data are limited or unavailable.

1. Introduction

As intelligent manufacturing becomes increasingly automated, digitized, and intelligent, more attention is paid to manufacturing process reliability and safety [1–3]. Minor failures in the production process can cause irreparable damage. Therefore, fault diagnosis is an important aspect of intelligent manufacturing [4–6]. Among the current machine learning methods, support vector machines, decision trees, and fractional order have been successfully applied in the field of fault diagnosis [7, 8]. In recent years, deep learning methods such as Bayesian networks, long short-term memory, and convolutional neural networks have been very popular in fault diagnosis due to their powerful modeling and representation capabilities [9–12]. The methods mentioned above can significantly improve the accuracy and efficiency of fault diagnosis under certain conditions. However, in order to obtain high fault diagnosis accuracy for deep learning methods, the primary condition

is that the source domain data should be sufficient and contain comprehensive fault diagnosis information. In practical industrial applications, machines are often in a dynamically changing production environment, and the health and fault information collected at this time are uncertain. Therefore, in the dynamically changing production process, it is difficult to collect a large amount of labeled fault data [13]. In addition, the machine is time-consuming and laborious to complete the degradation process, and the cost of marking a large amount of fault data is high. In order to prevent catastrophic accidents, many enterprises and factories do not allow machines to run to failure. Therefore, the above intelligent fault diagnosis methods are difficult to play a role in the dynamic changing production process [14].

In order to solve the above problems of insufficient training data and incomplete diagnostic information, some scholars have thought of transfer learning methods. This method can transfer a large amount of diagnostic information collected on a specific experimental platform to

dynamically changing production scenarios, solving the problem of insufficient training data [15]. For example, Zhang et al. proposed a first-layer wide convolutional deep neural network (WDCNN), the key of which is to use large convolution kernels in the first layer of convolution to extract short-term features. The convolution kernel parameters of the remaining convolutional layers except the first layer are reduced, which is conducive to deepening the network and suppressing overfitting [16]. Zhang et al. proposed a new CNN model, the advantage of which is that it does not require signal denoising preprocessing, which can realize fault diagnosis in noisy environments and variable working conditions [17]. Ren et al. proposed a new fault detection and classification method (DRCNN), which designed an important module “multiscale summation” for deep feature extraction. This method can combine features of multiple scales and different levels from unequal layers, which ensures the completeness of information [18]. However, the robustness of the diagnostic performance of transfer learning suffers in scenarios where working conditions and system characteristics are not fixed. Parameter transfer learning methods can assume that some parameters are shared between source tasks and target tasks, or the prior distribution of model hyperparameters is shared. Then, we use a small number of samples in the target domain to fine-tune the pretrained model, improve the overall performance of the model, and achieve a more robust fault diagnosis effect. However, the parameter transfer learning method also faces the problem of incomplete diagnostic information in the source domain [19].

With the rapid development and application of information technology, in recent years, digital twin (DT) technology has received more and more attention in various fields, such as product design and manufacturing, medical analysis, engineering construction, process optimization, and job shop scheduling. [20–22]. Also in the field of intelligent manufacturing, DT technology has also played a pivotal role and has become a powerful weapon to promote the development of intelligent manufacturing. It makes full use of physical model, sensor update, operation history, and other data, integrates multidisciplinary, multiphysical, multiscale, multiprobability simulation process, and completes the mapping in virtual space, thereby reflecting the full life cycle process of the corresponding physical equipment. DT technology can not only reduce design and maintenance costs but also improve manufacturing efficiency and quality. Ultrahigh-fidelity DT models can generate simulated data close to real systems, providing new opportunities for intelligent fault diagnosis. Wang et al. proposed a DT reference model for rotor system fault diagnosis. The requirements for building a digital twin model are discussed, and a model update scheme based on parameter sensitivity analysis is proposed to improve the adaptability of the model [23]. Jain et al. constructed a digital twin that can estimate the measurable characteristic output of a photovoltaic energy conversion unit (PVECU) in real time. A PVECU consists of a photovoltaic source and a source-level power converter [24]. Qin et al. proposed a full life cycle rolling bearing DT model driven by a combination of data and models. Through

an improved CycleGAN neural network, the simulated data in the virtual space are mapped to the physical space, and the results of the DT model are compared with the measured signals in the time and frequency domains to verify the effectiveness and feasibility of the proposed model [25]. Xu et al. proposed a two-stage DT fault diagnosis method (DFDD) based on deep transfer learning, which realizes fault diagnosis in the development and maintenance stages [26]. Qin et al. proposed a digital twin convolutional neural network model with multidomain input (DTCNNMI) in order to realize the misfire diagnosis of the diesel engine in a strong noise environment and different operating conditions [27]. However, most of the current research focuses on the conceptual model and key technologies of DT, and few people conduct more specific research on the fault diagnosis framework, mechanism, and algorithm to overcome the practical problem of limited diagnostic data.

In this paper, a novel digital twin-assisted fault diagnosis method for denoising autoencoder is proposed for the problem of machine intelligence fault diagnosis. The DT model can simulate a dynamically changing production process, thereby generating data of different fault states, and solving the problem of limited or unavailable fault state data for machines under dynamically variable production conditions. The main contributions of this paper are as follows:

- (1) A DT-assisted deep transfer learning fault diagnosis method is proposed, which is mainly used for fault diagnosis experiments of triplex pumps. A DT model of the machine is established to simulate the dynamically changing production process, thereby generating data for different failure states. The DT model is continuously updated during this process. The method solves the problem that the fault state data are limited or not used when the working state of the machine changes and the system characteristic changes.
- (2) A novel denoising autoencoder (NDAE) with Mish as the activation function is proposed, which has the properties of no upper bound, lower bound, smoothness, and nonmonotonicity compared with other activation functions.
- (3) A sparse penalty term is introduced to fully combine the advantages of sparse autoencoders and denoising autoencoders to effectively learn sparse feature representations from noisy samples.

2. Theoretical Background of Autoencoders

Convolutional neural network (CNN) is a commonly used network structure in deep learning methods and is currently widely used in the field of intelligent fault diagnosis of mechanical systems. However, the structure of CNN is relatively complex, and the amount of computation is relatively large compared with other deep learning methods. Compared with CNN, autoencoder has a simpler structure and stronger operability, which can train the model more easily and effectively. A type of neural network, after training, attempts to copy the input to the output. At

present, many improved forms have been derived from the autoencoder. On the basis of the autoencoder, the noise reduction autoencoder adds noise to the input data of the input layer in order to prevent the overfitting problem, so that the learned encoder is more robust. A sparse autoencoder is a special three-layer neural network with sparse constraints added to the general neural network. From the input layer to the hidden layer, the high-dimensional data are mapped to the low-dimensional data, and the projected low-dimensional data are restored to the original high-dimensional data from the hidden layer to the output layer [28]. Assuming that $x = [x_1, x_2, \dots, x_m]$ is a labeled m -dimensional real sample, the formula for noise sample $\tilde{x} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_m]$ is defined as follows:

$$\tilde{x} = x + N(0, \delta^2 \mathbf{I}), \quad (1)$$

where $N(0, \delta^2 \mathbf{I})$ represents Gaussian noise with noise level δ .

Then, the formulas of the feature vectors $\tilde{h} = [\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_n]$ and reconstruction vectors $\tilde{z} = [\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_m]$ of the noise samples are as follows:

$$\begin{aligned} \tilde{h} &= f_H(\mathbf{w}^{(1)}\tilde{x} + \mathbf{b}^{(1)}), \\ \tilde{z} &= f_O(\mathbf{w}^{(2)}\tilde{h} + \mathbf{b}^{(2)}), \end{aligned} \quad (2)$$

where f_H and f_O are the activation functions of the hidden layer and the output layer. $(\mathbf{w}^{(1)}, \mathbf{b}^{(1)})$ is the weights and biases of the input and hidden layers. Similarly, $(\mathbf{w}^{(2)}, \mathbf{b}^{(2)})$ is the weight and bias of the hidden layer and the output layer.

The formulas of the loss function l_1 , sparse penalty term l_2 , and weight decay term l_3 of MSE are expressed as follows:

$$\begin{aligned} l_1 &= \frac{1}{2} \sum_{i=1}^m (\tilde{z}_i - x_i)^2, \\ l_2 &= \beta \left(\sum_{j=1}^n r \log \frac{r}{\tilde{r}_j} + (1-r) \log \frac{1-r}{1-\tilde{r}_j} \right), \\ l_3 &= \frac{\lambda}{2} \left(\sum_{i=1}^m \sum_{j=1}^n \left((w_{ji}^{(1)})^2 + (w_{ji}^{(2)})^2 \right) \right), \end{aligned} \quad (3)$$

where β is the sparse penalty factor. r is the sparse constant. λ is the weight decay factor. $w_{ji}^{(1)}$ is the connection weight between the i th input unit and the j th hidden unit. Similarly, $w_{ji}^{(2)}$ is the connection between the j th hidden unit and the i th output unit connection weight.

Then, the overall loss function can be expressed as follows:

$$(L_S = l_1 + l_2 + l_3). \quad (4)$$

3. Proposed Method

3.1. NDAE Method. The NDAE method proposed in this paper is inspired by reference [28]. Based on the combination of sparse autoencoder and denoising autoencoder into sparse denoising autoencoder, a sparse penalty term is introduced, which can effectively learn the sparse features of

noise samples. In addition, inspired by reference [29], the Mish activation function with stronger learning ability is adopted. The ReLU activation function is the most widely used activation function in neural networks, and it mainly has the characteristics of having no upper bound and having a lower bound, which greatly limits its learning ability. Compared with ReLU, the Mish activation function has the characteristics of smoothness and nonmonotonicity. The smooth characteristics can make the network easier to optimize and improve the generalization performance, and the nonmonotonicity characteristics can improve the interpretability of the network. Comparison results on several datasets verify that Mish's metrics outperform ReLU and other activation functions for most tasks [29]. The waveform of the Mish function is shown in Figure 1. It can be seen from the figure that it allows a small negative gradient to flow in when it is negative, thereby ensuring information transfer and eliminating the dying ReLU phenomenon. The mathematical expression of the Mish activation function is as follows:

$$\begin{aligned} f_M(x) &= x \tanh(\text{softplus}(x)) \\ &= x \tanh(\ln(1 + e^x)), \\ \tanh(x) &= \frac{(e^x - e^{-x})}{(e^x + e^{-x})}, \end{aligned} \quad (5)$$

$$\text{softplus}(x) = \log(1 + e^x).$$

Mish is unbounded above and bounded below, and its first derivative can be defined as follows:

$$f'(x) = \frac{e^x \omega}{\delta^2}, \quad (6)$$

where $(\omega = 4(x+1) + 4e^{2x} + e^{3x} + e^x(4x+6))$, $(\delta = 2e^x + e^{2x} + 2)$.

The Mish activation function adds smoothness and nonmonotonicity to the ReLU activation function. These features can effectively retain the negative information of the data, make up for the deficiencies of ReLU, help information transfer, and have better expressiveness. It can be seen from formula (6) that the first derivative of the Mish activation function is differentiable; that is, the Mish activation function is continuously differentiable. This feature avoids singularities and thus avoids unwanted side effects when performing gradient-based optimization problems using the Mish activation function. In order to make the reconstructed output of the NDAE method infinitely close to the original input, a sigmoid activation function is selected at the output layer to normalize the input to the range of $[0, 1]$ into account. So the hidden and reconstructed outputs using the Mish activation function are

$$\begin{aligned} \tilde{h} &= f_M(\mathbf{w}^{(1)}\tilde{x} + \mathbf{b}^{(1)}), \\ \tilde{z} &= f_S(\mathbf{w}^{(2)}\tilde{h} + \mathbf{b}^{(2)}), \end{aligned} \quad (7)$$

where f_M and f_S are the Mish and sigmoid activation functions, respectively.

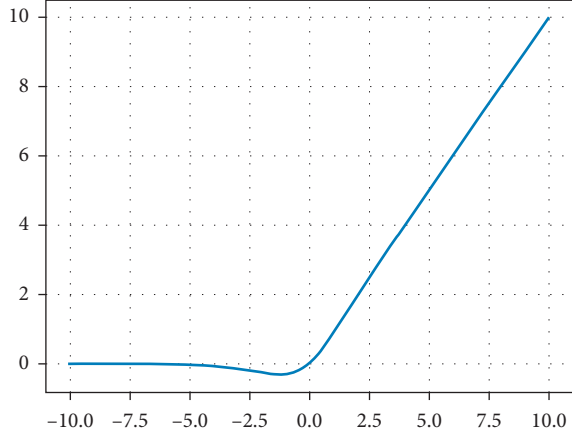


FIGURE 1: Mish activation function waveform.

To ensure that the real samples and the samples generated by the NDAE reconstruction are as similar as possible, the difference between the real samples and the reconstructed samples is reduced. To further measure the local similarity between the two, use the maximum correlation entropy instead of MSE, and use the gradient descent algorithm to adjust \mathbf{w} and \mathbf{b} , the correlation formula is as follows:

$$w_{q+1} = w_q - \xi_q \left(\frac{\partial L_N}{\partial w_q} \right) + \varepsilon (w_q - w_{q-1}),$$

$$b_{q+1} = b_q - \xi_q \left(\frac{\partial L_N}{\partial b_q} \right) + \varepsilon (b_q - b_{q-1}),$$
(8)

where q is the current number of iterations. L_N is the total loss function of the proposed method NDAE. ξ_q is the current learning rate. ε is the momentum factor. Then, the NDAE total loss function is as follows:

$$(L_N = -l_4 + l_2 + l_3),$$
(9)

where l_4 is the formula of maximum correlation entropy, which is more effective in local similarity measurement of complex signals than MSE. The formula for l_4 is as follows:

$$l_4 = \frac{1}{\sqrt{2\pi\tau}} \sum_{i=1}^m \exp\left(-\frac{(\tilde{z}_i - x_i)^2}{2\tau^2}\right),$$

$$\xi_{q+1} = \frac{\xi_q}{\rho},$$
(10)

where τ is the kernel width adjustment parameter. ρ is the decay factor. Using multiple NDAEs with softmax classifiers can be constructed to stack NDAEs to improve learning ability.

3.2. DT-Assisted NDAE Method. To solve the problem of limited or unavailable fault state data for machines under dynamically variable production conditions, a DT-assisted NDAE method is proposed in this paper. The overall framework of the method is shown in Figure 2. The green

part is the construction part of the DT model of the triplex pump. First, the simulation model of the real machine needs to be established, and then, the simulation model needs to be continuously updated to adapt to the dynamic and variable production environment. This paper updates the simulation model by minimizing the difference in system response between the simulation model and the measured data. The adaptively updated DT model is then used to simulate the fault state of the machine, generating comprehensive fault data required for fault diagnosis. The blue part in Figure 2 is the parameter transfer learning part of the new denoising autoencoder. First, the stacked NDAE model is constructed using the Mish activation function and maximum correlation entropy in 3.1, and then, a large amount of fault state data generated by the DT model are used as the training data in the source domain, which is input into the stacked NDAE model for pretraining. Finally, parameter transfer learning can greatly improve the training efficiency of stacked NDAE, so the parameter transfer learning method is used to realize machine fault diagnosis. It is worth noting that this paper selects a sample in the target domain to fine-tune the pre-trained stacked NDAE to further adjust the model parameters.

The shared parameters for parameter transfer learning in this paper are all hyperparameters, weights, and biases. It is worth noting that all weights and biases are pretrained before fine-tuning to ensure the effectiveness of parameter transfer learning.

4. Case Analysis

4.1. Experimental Description. In order to evaluate the effectiveness and feasibility of the proposed method, the method is applied to a fault diagnosis example of a triplex pump. The DT model of the triplex pump is shown in Figure 3. Inspired by reference [30], this paper imitates reference [30] and uses the Simscape module in Matlab to create a simulation model of a triplex pump. Triplex pumps have a crankshaft driving three plungers. Compared to single-piston pumps, one air chamber of the plunger is always vented, resulting in smoother flow and less pressure variation, thereby reducing material strain. The parameter values were then automatically tuned using Simulink design optimization so that the model produced results that matched the measured data to simulate the system behavior of a triplex pump in a dynamically variable production environment. Simulink design optimization selects parameter values for simulation, calculates the difference between the simulation curve and the measured curve to update the simulation model, and generates a simulation model with the system response function of modifying model parameters. Based on this difference, new parameter values are selected for a new simulation. The gradient of the parameter value is calculated to determine the direction in which the parameter should be adjusted. The DT model update of the triplex pump in this study is implemented through Simulink design optimization and automatically tunes the parameter values so that the model generates results that match the measured data.

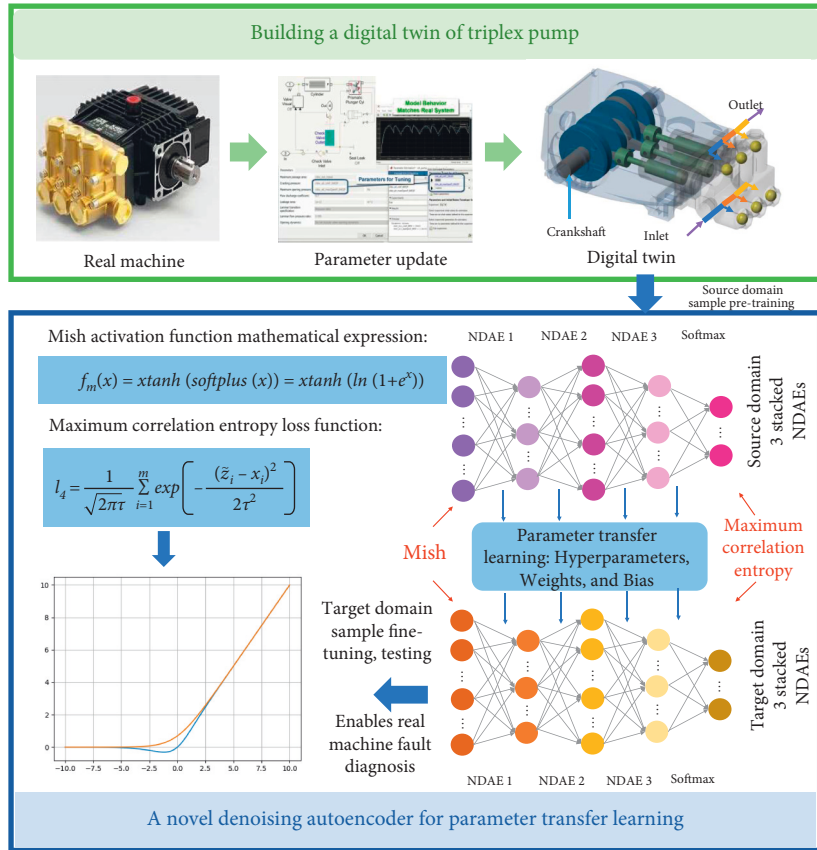


FIGURE 2: General framework of DT-assisted deep transfer learning method.

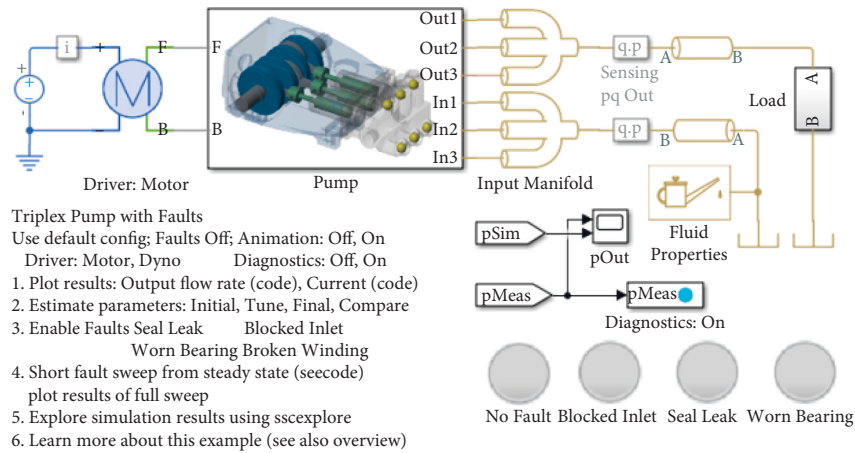


FIGURE 3: DT model of triplex pump.

The DT model of the triplex pump can simulate three typical pump failures, including seal leakage from the plunger, inlet blockage, and increased friction due to bearing wear. These fault conditions can be configured and toggled through the pump module dialog or commands. This paper collects data for seven fault states, including healthy state, three single faults, and three composite faults. Dataset A and dataset B are collected by simulating two scenarios. Dataset A is simulation data with original parameters. Dataset B is

simulation data collected when working conditions and system characteristics change, that is, actual situation data. Table 1 shows the detailed description of dataset A and dataset B, 125 samples are selected for each fault state, and each sample contains 1200 data points. Table 2 shows the detailed settings of the DT-assisted NDAE parameter transfer learning task. Table 3 shows the hyperparameter settings for stacked NDAE. The size of the first, second, and third hidden layers and other network structures are

TABLE 1: Seven working states of triplex pump.

Datasets	Working status	Number of samples	Labels
A/B	Healthy	125	1
	Seal leak	125	2
	Blocked inlet	125	3
	Bearing wear	125	4
	Seal leak and blocked inlet	125	5
	Seal leak and bearing wear	125	6
	Blocked inlet and bearing wear	125	7

TABLE 2: Detailed settings of parameter transfer tasks.

Method	Source domain training/test samples	Number of training/testing samples
Parameter transfer learning	Dataset A/B	75/50

TABLE 3: Hyperparameter settings for stacked NDAE methods.

Hyperparameters	Value	Hyperparameters	Value
The size of the first hidden layer	450	Kernel width	1.2
The size of the third hidden layer	100	Sparse penalty factor	5
Number of iterations	60	Initial learning rate	0.01
Weight decay coefficient	0.004	Decay factor	1.1
Noise level	0.08	Momentum	0.8

determined by experiments and experience. The number of iterations, initial learning rate, decay factor, and momentum are determined empirically, respectively. The selection of other hyperparameters is mainly based on reference [25].

4.2. Comparison Method. To verify the effectiveness and feasibility of the proposed method, it is compared with several state-of-the-art methods. Both the comparison method and the proposed method are tested using the fault data generated by the DT model.

- (1) SVM. The SVM algorithm is used to realize the fault classification of the triplex pump. SVM is a binary classification model that maps feature vectors to points in space, and its purpose is to find a line to better distinguish these points. Before the advent of deep learning, SVM was considered to be the better-performing algorithm in machine learning.
- (2) Stacked SDAE method. Stacked denoising autoencoders with ReLU as activation function.
- (3) LeNet-5 CNN. Fault classification of triplex pumps using a classic LeNet-5 convolutional neural network.

4.3. Experimental Results and Analysis. In this experiment, the effectiveness and feasibility of the proposed NDAE method are verified by the parameter transfer learning

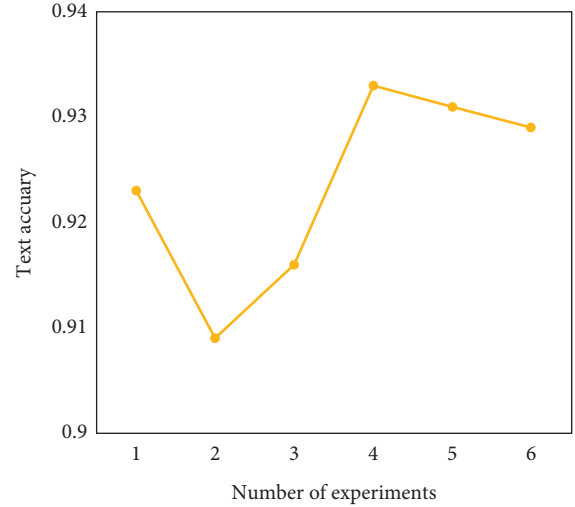


FIGURE 4: The fault diagnosis accuracy of the proposed method in ten experiments.

method. First, 75 samples are randomly selected from the 125 samples of dataset A as training samples in the source domain, and these 75 samples are input into the NDAE network for pretraining. Then, one sample is used from dataset B to fine-tune the pretrained network. This is because the DT model of the triplex pump can generate the data of the fault state, so just select a sample from the target domain to fine-tune the pretrained stacked NADE, and further construct the deep structure NADE to obtain better fault diagnosis results. Finally, 50 samples are used from dataset B for testing. In order to reduce the influence of random factors, the experiments were repeated six times; that is, six independent experiments were carried out using random samples for each method. The fault diagnosis accuracy of six experiments is shown in Figure 4. It can be seen from Figure 4 that the diagnostic accuracy of the six experiments exceeds 90%, and the average fault diagnosis accuracy is 92.4%.

To verify the effectiveness and feasibility of the proposed method, it is compared with SVM, stacked SDAE method, and LeNet-5 CNN method. To reduce the influence of random factors, the experiments were repeated six times; that is, six independent experiments were performed using random samples for each method. The experimental results are shown in Figure 5. The average accuracy of the SVM method, stacked SDAE method, and LeNet-5 CNN method is 76.5%, 87.8%, and 88.6%, respectively. It can be seen from the experimental results that the proposed method has higher fault diagnosis accuracy and is more conducive to the fault classification of the triplex pump.

5. Conclusion

In this paper, a DT-assisted NDAE parameter transfer learning fault diagnosis method is proposed, which is mainly used in the fault diagnosis experiment of the triplex pump.

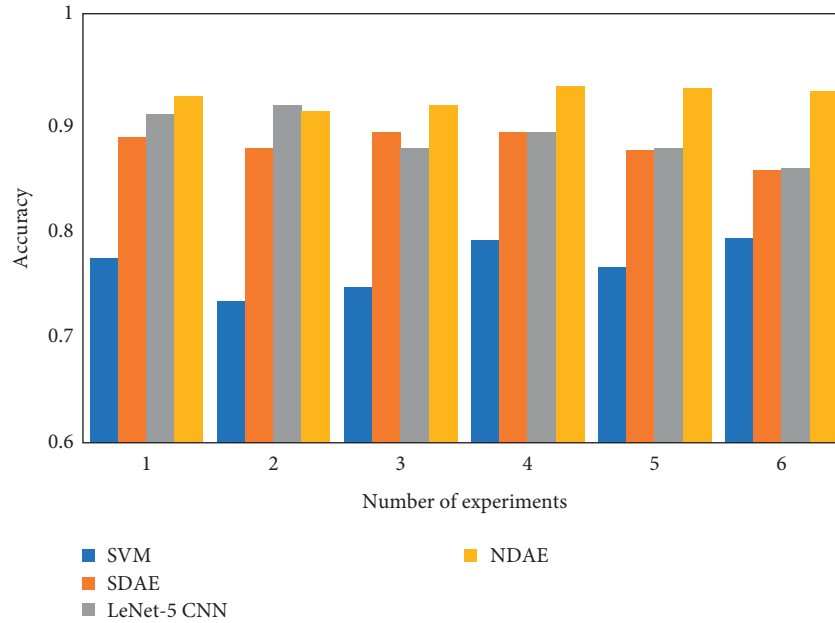


FIGURE 5: Fault diagnosis accuracy of four methods.

This method is designed to achieve high-accuracy fault classification when measurement data are insufficient or unavailable. We use the digital twin model of the machine to generate fault state data similar to the actual system to make up for the lack of data. In addition, the stacked autoencoder is improved, and the Mish activation function has no upper bound, lower bound, nonmonotonicity, and smoothness to increase the generalization performance of the network and the interpretability of the network. This ensures information transfer and eliminates the dying ReLU phenomenon. Finally, by generating simulation data of the triplex pump under various fault conditions, the effectiveness of fault diagnosis of the proposed NDAE method is verified. The results show that the ultrahigh-fidelity DT model can generate simulated data close to the real system, providing new opportunities for intelligent fault diagnosis. The DT-assisted NDAE parameter transfer learning fault diagnosis method can realize intelligent fault diagnosis of mechanical systems in dynamically changing production environments.

Although the DT-assisted NDAE parameter transfer learning fault diagnosis method can effectively improve the fault diagnosis accuracy of the model, the construction of the machine's DT model is a difficulty of this method. Therefore, the following research focuses on making full use of the main mechanism of DT to build DT models of other basic components such as bearings and combining deep transfer learning methods to improve fault diagnosis performance. How to further combine DT and deep transfer learning is the focus and difficulty of the next research.

Data Availability

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Wenan Cai wrote the paper; Qianqian Zhang revised the manuscript; Jie Cui revised the code of the manuscript. All authors have read and approved the final manuscript.

Acknowledgments

The authors would like to acknowledge Shanxi Province Scientific and technological innovation project of colleges and universities (2020L0606) and Shanxi Province Graduate Student Innovation Project (2021Y583).

References

- [1] A. Kusiak, "Smart manufacturing," *International Journal of Production Research*, vol. 56, no. 1-2, pp. 508–517, 2018.
- [2] S. Duan, W. Song, E. Zio, C. Cattani, and M. Li, "Product technical life prediction based on multi-modes and fractional Lévy stable motion," *Mechanical Systems and Signal Processing*, vol. 161, no. 5, Article ID 107974, 2021.
- [3] H. Liu, W. Song, Y. Zhang, and A. Kudreyko, "Generalized cauchy degradation model with long-range dependence and maximum lyapunov exponent for remaining useful life," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, Article ID 3512812, 2021.
- [4] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: a review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, Article ID 106587, 2020.
- [5] Y. Li, X. Liang, Y. Yang, M. Xu, and W. Huang, "Early fault diagnosis of rotating machinery by combining differential

- rational spline-based LMD and K-L divergence,” *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 11, pp. 3077–3090, 2017.
- [6] R. Liu, B. Yang, E. Zio, and X. Chen, “Artificial intelligence for fault diagnosis of rotating machinery: a review,” *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, Aug. 2018.
 - [7] Z. Yin and J. Hou, “Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes,” *Neurocomputing*, vol. 174, pp. 643–650, Jan. 2016.
 - [8] H. Wang, J. Long, Z. Liu, F. You, and V. Ponomaryov, “Fault characteristic extraction by fractional lower-order bispectrum methods,” *Mathematical Problems in Engineering*, vol. 2020, Article ID 8823389, 24 pages, 2020.
 - [9] L. Bennacer, Y. Amirat, A. Chibani, A. Mellouk, and L. Ciavaglia, “Self-diagnosis technique for virtual private networks combining bayesian networks and case-based reasoning,” *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 1, pp. 354–366, Jan. 2015.
 - [10] X. He, Z. Wang, and Y. Li, “Joint decision-making of parallel machine scheduling restricted in job-machine release time and preventive maintenance with remaining useful life constraints,” *Reliability Engineering & System Safety*, vol. 222, 2022.
 - [11] W. Zhao, Z. Wang, W. Cai et al., “Multiscale inverted residual convolutional neural network for intelligent diagnosis of bearings under variable load condition,” *Measurement*, vol. 188, Article ID 110511, 2022.
 - [12] Z. Wang, J. Cui, W. Cai, and Y. Li, “Partial transfer learning of multidiscriminator deep weighted adversarial network in cross-machine fault diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.
 - [13] J. Cui, Y. Li, Q. Zhang, Z. Wang, W. Du, and J. Wang, “Multi-layer adaptive convolutional neural network unsupervised domain adaptive bearing fault diagnosis method,” *Measurement Science and Technology*, vol. 33, no. 8, Article ID 085009, 2022.
 - [14] Z. Wang, N. Yang, and N. Li, “A New Fault Diagnosis Method Based on Adaptive Spectrum Mode Extraction,” *Structural Health Monitoring*, vol. 20, 2021.
 - [15] X. Li, Y. Hu, M. Li, and J. Zheng, “Fault diagnostics between different type of components: a transfer learning approach,” *Applied Soft Computing*, vol. 86, Article ID 105950, 2020.
 - [16] W. Zhang, W. Ouyang, and L. Wen, “Collaborative and adversarial network for unsupervised domain adaptation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
 - [17] Z. Pei, Z. Cao, M. Long, and J. Wang, “Multi-adversarial domain adaptation,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, February 2018.
 - [18] Z. Wang, J. Zhou, Y. DuLei, W. Lei, and J. Wang, “Bearing fault diagnosis method based on adaptive maximum cyclostationarity blind deconvolution,” *Mechanical Systems and Signal Processing*, vol. 162, Article ID 108018, 2022.
 - [19] R. Yan, F. Shen, C. Sun, and X. Chen, “Knowledge transfer for rotary machine fault diagnosis,” *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8374–8393, 2020.
 - [20] F. Tao, F. Sui, A. Liu et al., “Digital twin-driven product design framework,” *International Journal of Production Research*, vol. 57, no. 12, pp. 3935–3953, 2019.
 - [21] Y. Fang, C. Peng, P. Lou, Z. Zhou, J. Hu, and J. Yan, “Digital-twin-based job shop scheduling toward smart manufacturing,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 12, pp. 6425–6435, 2019.
 - [22] Y. Yi, Y. Yan, X. Liu, Z. Ni, J. Feng, and J. Liu, “Digital twin-based smart assembly process design and application framework for complex products and its case study,” *Journal of Manufacturing Systems*, vol. 58, 2020.
 - [23] J. Wang, L. Ye, R. X. Gao, C. Li, and L. Zhang, “Digital Twin for rotating machinery fault diagnosis in smart manufacturing,” *International Journal of Production Research*, vol. 57, no. 12, pp. 3920–3934, 2019.
 - [24] P. Jain, J. Poon, J. P. Singh, C. Spanos, S. R. Sanders, and S. K. Panda, “A digital twin approach for fault diagnosis in distributed photovoltaic systems,” *IEEE Transactions on Power Electronics*, vol. 35, no. 1, pp. 940–956, 2020.
 - [25] Y. Qin, X. Wu, and J. Luo, “Data-model combined driven digital twin of life-cycle rolling bearing,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1530–1540, 2022.
 - [26] Y. Xu, Y. Sun, X. Liu, and Y. Zheng, “A digital-twin-assisted fault diagnosis using deep transfer learning,” *IEEE Access*, vol. 7, pp. 19990–19999, 2019.
 - [27] C. Qin, Y. Jin, J. Tao, D. Xiao, H. Yu, and C. Liu, “DTCNNMI: A deep twin convolutional neural networks with multi-domain inputs for strongly noisy diesel engine misfire detection,” *Measurement*, vol. 180, 2021.
 - [28] M. Xia, H. Shao, D. Williams, and L. Shu, “Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning,” *Reliability Engineering & System Safety*, vol. 215, 2021.
 - [29] M. Diganta, “Mish: a self regularized non-monotonic activation function,” *Machine learning*, vol. 1, 2019.
 - [30] S. Miller, “Predictive maintenance using a digital twin,” 2020, <https://uk.mathworks.com/company/newsletters/articles/predictive-maintenanceusing-a-digital-twin.html>.

Research Article

MFA: A Smart Glove with Multimodal Intent Sensing Capability

Hongyue Wang,^{1,2} Zhiquan Feng ,^{1,2} Jinglan Tian,^{1,2} and Xue Fan^{1,2}

¹School of Information Science and Engineering, University of Jinan, Jinan 250022, China

²Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan 250022, China

Correspondence should be addressed to Zhiquan Feng; ise_fengzq@ujn.edu.cn

Received 23 April 2022; Revised 15 June 2022; Accepted 20 June 2022; Published 11 July 2022

Academic Editor: Zhongxu Hu

Copyright © 2022 Hongyue Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

At present, virtual-reality fusion smart experiments mostly employ visual perception devices to collect user behavior data, but this method faces the obstacles of distance, angle, occlusion, light, and a variety of other factors of indoor interactive input devices. Moreover, the essence of the traditional multimodal fusion algorithm (TMFA) is to analyze the user's experimental intent serially using single-mode information, which cannot fully utilize the intent information of each mode. Therefore, this paper designs a multimodal fusion algorithm (hereinafter referred to as MFA, Algorithm 4) which focuses on the parallel fusion of user's experimental intent. The essence of the MFA is the fusion of multimodal intent probability. At the same time, this paper designs a smart glove based on the virtual-reality fusion experiments, which can integrate multichannel data such as voice, visual, and sensor. This smart glove can not only capture user's experimental intent but also navigate, guide, or warn user's operation behaviors, and it has stronger perception capabilities compared to any other data glove or smart experimental device. The experimental results demonstrate that the smart glove presented in this paper can be widely employed in the chemical experiment teaching based on virtual-reality fusion.

1. Introduction

Many secondary school chemistry experiments currently contain damaging, costly, and sometimes dangerous experimental qualities, preventing students from doing practical operations for many of them. Simultaneously, because teachers have a finite amount of teaching energy, some students are prone to overlooking the most important aspects of experimental operation and exhibiting irregular operation behavior when conducting experiments. As a result, developing an intelligent, operable, and low-risk experimental platform system for secondary school experiments has become a pressing issue.

Three popular and effective technical solutions are available. The first is to create a virtual experimental platform [1] using virtual modeling software, which solves the problem of high costs and risks associated with teaching real experiments. However, most existing virtual experimental platforms use a mouse, a keyboard, and a monitor for

experiments, reducing the user's sense of operation significantly. The second is the use of virtual-reality (VR) and augmented reality (AR) technologies, which allows people to have the same experience as if they were participating in a real experiment but require more memory. The third is to build a virtual-reality fusion experimental platform [2]. Because the existing platform relies on several cameras or indoor interactive input devices like KINECT to monitor the user's actions, issues like occlusion and the inability to observe minor experimental phenomena are common.

In light of the shortcomings of the previous three technical solutions, this paper finds that just one or two sensors can be employed to detect the user's hand position and movement in the study of the smart glove. As a result, this paper designs a smart glove and a mixed reality (MR) experiment system that can fuse multimodal data and, using a multimodal fusion algorithm, determine the user's experimental intent and direct the experiment, as well as correct and remind the user of any incorrect or unsafe

procedures. The experimental intent of this paper refers to the experimental steps of user operation. By obtaining the real-time experimental intent of the user, the smart glove system can advise the user in time or correct and remind the user of incorrect and harmful steps. And the foundation for completing the experiment is understanding the user's experiment intent.

Therefore, the following are the primary innovations in this paper:

- (1) This paper designs a smart glove system based on MR experiment teaching, to overcome the limits of existing interaction devices and tools. Using multi-sensing fusion technology, the smart glove can complete complex user experiment intent analysis using only some simple sensors. In this study, a monocular camera is integrated at the wrist of the smart glove to solve the problems of visual occlusion and observation of subtle phenomena. Not only can the device capture data about the user's behavior throughout the experiment, but it can also detect the user's experimental intent and demand for the experimental scene.
- (2) Under the background of intelligent experiment, this paper proposes a multimodal fusion algorithm (MFA) which fuses the user's experimental intention in parallel at the intention level, and solves the limitations of the traditional multimodal fusion algorithm (TMFA) which is used to analyzing the user's behavior in serial. After acquiring the intent probability in the user's voice, visual, and sensor channels, this paper employs information weight method to fuse the user intent probabilities. The essence of MFA is to convert the user's abstract intent which is difficult to understand into a calculable probability problem.

This paper is organized as follows. Section 2 comprehensively discusses related work. Section 3 describes the prototype design of smart glove and construction of MR virtual-reality fusion laboratory. Section 4 introduces the overall framework and MFA. Section 5 analyzes and discusses the experimental results. Conclusions are presented in Section 6.

2. Related Work

2.1. Experimental Teaching of Virtual-Reality Fusion. The virtual-reality integration experiment is an experiment in which users interact with the virtual world using real-world experimental objects to interact with the virtual objects in a computer-simulated virtual environment.

Virtual experiments were first proposed by William Wolfe in 1989, and as computer multimedia graphic picture technology advanced, virtual experiments were gradually included in educational training. At the beginning of the development of virtual experimental teaching mainly based on Web technology, Aljuhani et al. [3] developed a virtual laboratory platform based on the Web platform, and users can conduct virtual experiments using a mouse.

Virtual-reality technology has progressively become popular as research progresses. Bogusevschi et al. [4] used virtual-reality technology to recreate the water cycle system in nature, and students were able to observe and study according to the program's principles. Salinas and Pulido [5] used AR technology to create a virtual platform to help students better understand conic curves. With the advancement of AR technology, the use of MR technology for teaching and learning has entered our vision. MR is a computerized virtual-reality technology that allows the real and virtual worlds to be presented and interacted within the same visual area. Lenz et al. [6] designed an MR speech lab that combines real and virtual classrooms, which can realistically portray the number of students and other noises that may be generated, while the teacher simulates daily teaching in a virtual environment using MR displays. Hu et al. [7] proposed a vision-based dynamic head posture tracking system, which improved the driving simulator's immersion and engagement.

In comparison with genuine trials, the virtual-reality fusion experiment not only increases users' interest in learning and aids their comprehension of knowledge, but also reduces consumables and risks. Virtual-reality fusion studies, as opposed to virtual experiments that divide the real from the virtual, provide an interactive feedback pathway between the virtual world, the real world, and the user, increasing the realism of the user experience. Researchers are gradually mixing MR technology with experimental education instruction, and MR technology has therefore become a popular study direction because it possesses the features of virtual-reality fusion and real-time interaction.

2.2. Understanding the Intent of Multimodal Fusion.

Originally, multimodal fusion meant combining various senses. The use of more than one input channel (e.g., gesture, speech, visual, haptic, etc.) to communicate with a machine in a system is referred to as multimodal interaction. As a result, adopting multimodal input interaction in virtual-reality settings might improve the naturalness and efficiency of interaction compared to using unimodal [8].

Ismail et al. [9] merged voice and gestures in the context of virtual-reality interaction, making the user's operation of dealing with virtual items in an AR environment more natural. Kadavasa and Oliver [10] created a virtual-reality driving system for autistic individuals that included physiological signals, brain signals, and eye gaze information, to improve autistic patients' driving abilities. Due to the lack of practical utility and popularity of virtual-reality experimental teaching, Xiao et al. [11] developed a multimodal interaction model that integrates voice and sensor information. Liu et al. [12] proposed a deep learning-based multimodal fusion model that combines three modal data sets: voice commands, hand gestures, and body movements, using various deep neural networks. In the field of automatic driving, Hu et al. [13] introduced contrastive learning approach to train a feature extractor with good representation ability in order to improve driving performance and avoid possible fatal accidents. In the field of speech recognition,

Ondas et al. [14] proposed a combination of modified LIMA framework and iterative spectral subtraction algorithm to improve the robustness of speech recognition in noisy environment.

In short, the multimodal fusion interactive approach can solve the problems of input incomprehension, incompleteness, and misunderstanding caused by unimodal interaction with the system, as well as the ambiguity caused by relying on only one modality's input information to understand the user's intent. Existing multimodal fusion algorithms, on the other hand, primarily use a variety of single-channel information to assess user intent serially, but they are unable to fuse multichannel information to analyze user intent in parallel.

2.3. Smart Glove. The interaction method of keyboard and mouse combination has the attribute of having a weak sense of genuine operation in the field of human-computer interaction. Using traditional indoor interactive input devices for behavior detection, on the other hand, will result in the problems of occlusion and inability to observe subtle experimental phenomena. As a result, the above two human-computer interface methods each have their drawbacks. However, the advent of the smart glove gives a more natural human-computer interaction tool. It can interact with real objects using its presence of sensing devices and is not limited by the camera's field of vision, and it has responsiveness, good real-time, and high precision. Therefore, the smart glove is widely employed in a variety of applications, including sign language recognition, robot manipulation, and rehabilitation training.

Lokhande et al. [15] developed a glove that uses flex sensors and attitude sensors to transform gestures into text format in the field of conducting sign language recognition. Abhijith Bhaskaran et al. [16] created a smart glove that can recognize behaviors, gather hand position data, and transform recognized sign language into speech broadcast. Kumar Mummadi et al. [17] integrated an IMU module inside the glove's fingers, and the experimenter detected the wearer's hand posture and motion trajectory to perform French sign language recognition.

In the realm of robot manipulation and rehabilitation, the smart glove has also played an essential role. To enable remote manipulation of a robot, Roy et al. [18] designed a glove with flex sensors. Ma et al. [19] developed a smart glove to assist patients with weak hands in performing certain actions based on the direction and strength of fingertip movements during the rehabilitation process. Liu et al. [20] created a smart glove that uses inertial and magnetic measurement unit sensors to reconstruct hand movements accurately. Ge et al. [21] created a data glove that uses flex sensors to forecast the final gesture at the conclusion of the user's hand motion in real time.

In conclusion, the current virtual experiment platform primarily completes experiments by playing virtual animations, and users lack real-world experience with the platform. Simultaneously, other experiments rely primarily on unimodal interaction, which lacks an effective

understanding and feedback of user intent. Although a traditional data glove can collect user data, it lacks knowledge of the user's intent and perception of external information during human-computer contact, and most of them only employ a single channel for interaction. Therefore, this paper proposes a smart glove with cognitive capabilities based on multimodal fusion, in which a multimodal fusion algorithm fuses user intent from visual, sensor, and voice channels in parallel at the intention level. The smart glove system can also employ MR technology to show the corresponding experimental phenomena and achieve the functions of guidance and error correction for the user after acquiring the user's final experimental intent.

3. Design of Virtual-Reality Fusion Experimental System Based on Smart Glove

3.1. Hardware Equipment. This paper designs a new smart glove with multisensing fusion in the context of a secondary school experiment. This smart glove can collect multimodal data about the user and then detect the user's operational intent, which is characterized by convenience, operability, and flexibility. Figure 1 depicts a physical prototype of the smart glove.

Flex sensors, attitude sensors, pressure sensors, vibration motor modules, a monocular camera in the wrist part of the glove, and an indoor binocular camera applied to an MR experimental system make up the hardware element of the smart glove system. Different sensors can gather different information about the user's hands. To improve the traditional data glove which seriously affects the user's operation due to the complicated wires, this paper improves the data transmission between the smart glove and the computer to Bluetooth wireless transmission mode. The following is the functional design of each of its components.

- (1) Flex sensor: This sensor (Flex Sensor 4.5) is positioned in the smart glove's finger section and is used to obtain the degree of bending of the user's finger. The smart glove can map the bending changes of the sensor to the virtual hand in the Unity virtual scene, which is then utilized to restore the user's finger's bending state in real time.
- (2) Attitude sensor: This sensor (MPU6050) is positioned in the back of the smart glove's hand and is used to restore the user's hand's real-time posture by measuring the three-axis angle, velocity, and acceleration of the user's hand movement.
- (3) Pressure sensor: This sensor is located at the end of the smart glove's finger and is responsible for monitoring the user's finger end movements during operation.
- (4) Vibration motor module: The sensor is positioned in the back part of the smart glove's hand, and it can provide vibration feedback when the user grasps a real or virtual object.
- (5) Monocular camera: The camera is fixed in the wrist part of the smart glove, which can obtain the

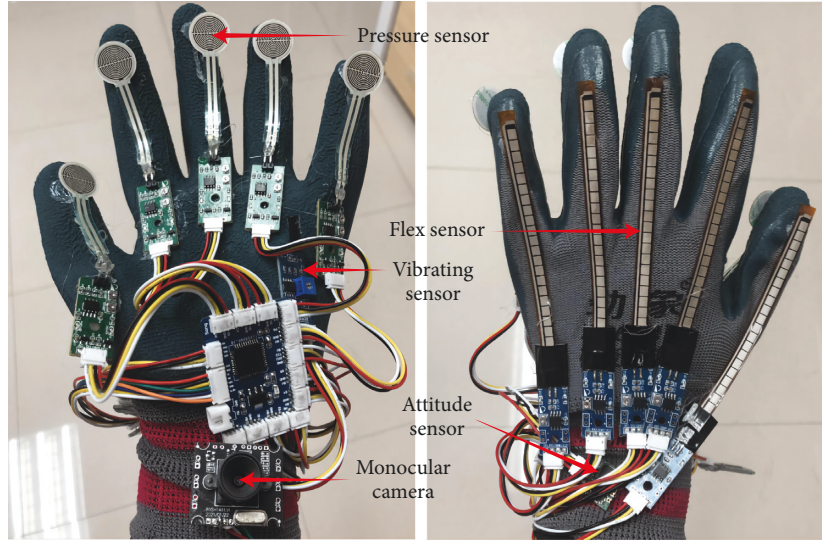


FIGURE 1: A physical prototype of the smart glove.

information of objects in experimental scene in real time and observe the subtle experimental phenomena. Its appearance solves the occlusion problem of traditional indoor input devices when identifying objects.

- (6) Indoor binocular camera for virtual-reality fusion: This camera is different from the occlusion-causing indoor camera. This binocular camera and Unity's Vuforia plug-in are used by the MR experimental system to provide a conduit between the actual world and the virtual world, allowing real experimental objects to interact with virtual experimental objects. Simultaneously, it can track the smart glove's real-time motion trajectory.

Trajectory Acquisition of Smart Glove Based on SGBM and YOLOv5 Fusion

During the experiment, the smart glove needs to acquire the user's hand movement trajectory and map it to the virtual experiment scene built by Unity in real time. The purpose is to enable the smart glove to interact with the experimental objects in the virtual-reality fusion experimental scene, which is the foundation for determining the user's experimental intent.

This paper incorporates the indoor binocular camera's binocular range function into the YOLOv5 target recognition procedure, while the smart glove is moving. The depth coordinates of the smart glove are obtained using the binocular stereo matching algorithm SGBM, with camera calibration, stereo correction, stereo matching, and range as the primary experimental steps. This fusion technique allows YOLOv5 to obtain not only the smart glove's position in the plane direction (i.e., x -axis and y -axis), but also the smart glove's depth value z , which is one of the paper's more difficult issues.

In this paper, the position of the indoor camera is used as the coordinate origin to obtain the 3D position of the smart glove in real time, as shown in Figure 2.

The coordinate information is processed using equation (1) based on the coordinate mapping relationship between the virtual scene and the camera position once the user's hand movement trajectory is collected. It is worth mentioning that (Pos_x, Pos_y, Pos_z) is the smart glove's 3D position mapped to the virtual world, and k is the coordinate transformation's scale factor. Finally, the data are transferred to the Unity platform through a socket connection.

$$\begin{bmatrix} Pos_x \\ Pos_y \\ Pos_z \end{bmatrix} = k \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \quad (1)$$

3.2. MR Experiment Scene Building. The majority of virtual laboratories are provided in the form of augmented reality (AR) or virtual-reality (VR), with participants immersed in a purely imaginary scenario of the experiment without being able to observe their actions in detail. As a result, this paper uses Unity's Vuforia plug-in and an indoor binocular camera to create an MR lab. This lab can integrate virtual sceneries and the actual world using AR technology, allowing them to cohabit and interact. MR is divided into two components in its practical application. The virtual object exhibited outside the screen is one component, and the real scene displayed inside the screen is the other.

The footage from the indoor camera is combined with the virtual world, creating a new visualization of the experimental environment. This fusion process allows users in the MR lab to genuinely watch their operations as well as observe the experimental phenomena that arise as a result of those operations as shown in Figure 3.

The smart glove system incorporates a speech recognition module to better conduct the virtual-reality fusion experiment teaching and collect the user's intent. The Baidu voice SDK is used in this study to continually monitor and recognize the user's voice information. And the MFA is used

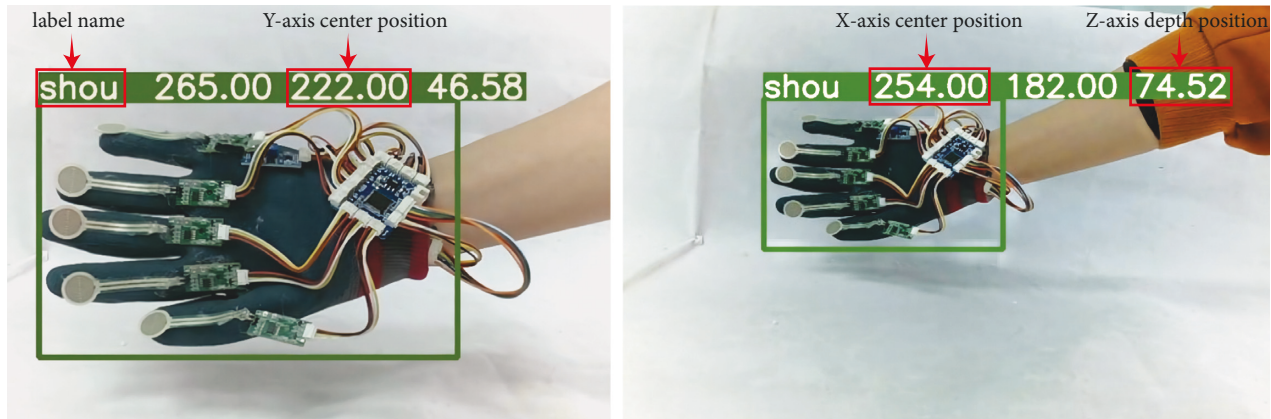


FIGURE 2: The three-dimensional position of the glove comparison picture (far and near position).



FIGURE 3: Comparison between traditional AR chemistry laboratory and MR chemistry laboratory.

to calculate the intention probability after the effective information in the experiment is obtained; at the same time, the module uses the Unity's SpVoice plug-in to provide users with voice guidance or explain experimental phenomena, reducing the user's memory burden.

4. Design of Intelligent Cognitive Module

4.1. Overall Framework. To guide and correct users' experimental operations, this paper uses the cognitive-behavioral theory [22], which refers to altering bad cognition by modifying the user's thinking and behavior. Users' thinking and actions should be tightly interwoven in the ideal situation. However, in real-world situations, users' cognition formation will be influenced by their automatic thinking; i.e., some users' actions will perform certain incorrect operations without thinking about it. As a result, the development of cognitive behavior theory can assist users in rationally correcting their unthinking conduct. Based on this, this paper constructs a general framework of a smart glove system based on multimodal fusion, as shown in Figure 4.

In the input layer, the smart glove acquires multimodal information from the user's voice, visual, and sensor channels, and passes the data into the recognition layer. The recognition layer and the fusion layer are critical components of the smart glove's overall framework. Under the background of intelligent experiment, the TMFA has the flaw of user experiment intention obtained by serial fusion of single-mode information. To overcome this challenge, the intent probability models for the voice, visual, and sensor channels in the recognition layer are

established, and they are utilized to update the intent probability of their respective channels in real time. By studying user behavior, these intent probability models convert the abstract intent of user behavior under each channel into a computable set of intent probabilities by analyzing the user behavior. In the fusion layer, this study ingeniously proposes the MFA, which dynamically updates the associated weight of each channel using the information weight method, fuses the intent probability set produced by the recognition layer in parallel, and finally determines the user's experimental intent.

In summary, the essence of the model and algorithm proposed in this paper is to convert the abstract intent, which is difficult to speculate, into a mathematical language that can be computed.

4.2. Model of Visual Channel Based on YOLOv5. The monocular camera on the smart glove can perceive the entire experimental scenario through the visual channel. To obtain the user's experimental intent under the visual channel, this model uses the incremental change of the bounding box area of the experimental object in YOLOv5 to infer the probability of the user's experimental intent. The target object obj_i will be dynamically updated each time the user performs an experimental operation. The visual channel intent probability acquisition algorithm (hereinafter referred to as VSIPAA) is based on YOLOv5, as shown in Algorithm 1.

Different experimental items correspond to one or more experimental intent in the visual channel. When the VEIPAA algorithm updates the usage probability of the

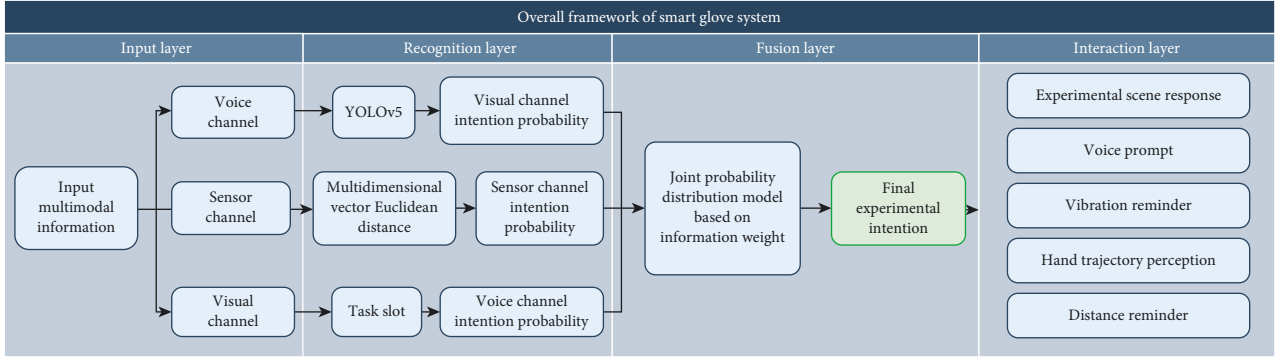


FIGURE 4: The overall framework of smart glove system based on multimodal fusion.

Input: Collection of experimental objects captured by the smart glove's monocular camera \mathbf{OBJ} .

Output: Experimental intent probability set \mathbf{Visual} under the visual channel.

- (1) **While** $\mathbf{OBJ} \neq \mathbf{Empty}$ **do**
- (2) YOLOv5 obtains the output coordinates of the bounding box of object \mathbf{obj}_i in the object set \mathbf{OBJ} at frame \mathbf{t} of the smart glove, $[\mathbf{x}(\mathbf{t})_{\min}, \mathbf{y}(\mathbf{t})_{\min}, \mathbf{x}(\mathbf{t})_{\max}, \mathbf{y}(\mathbf{t})_{\max}] = \mathbf{getPosition}((\mathbf{obj}_i, \mathbf{t}))$
- (3) When the smart glove moves, YOLOv5 gets the output coordinates of the bounding box of object \mathbf{obj}_i in the object collection \mathbf{OBJ} at frame $\mathbf{t} + 1$, $[\mathbf{x}(\mathbf{t} + 1)_{\min}, \mathbf{y}(\mathbf{t} + 1)_{\min}, \mathbf{x}(\mathbf{t} + 1)_{\max}, \mathbf{y}(\mathbf{t} + 1)_{\max}] = \mathbf{getPosition}((\mathbf{obj}_i, \mathbf{t} + 1))$
- (4) Calculate the area of object \mathbf{obj}_i at frame \mathbf{t} , $\mathbf{S}(\mathbf{t})_i = \mathbf{getArea}(\mathbf{obj}_i, \mathbf{t})$
- (5) Calculate the area increment of object \mathbf{obj}_i in the object set \mathbf{OBJ} between two frames, $\mathbf{S}(\mathbf{t}, \mathbf{t} + 1)_i = \mathbf{getArea}(\mathbf{obj}_i, \mathbf{t} + 1) - \mathbf{getArea}(\mathbf{obj}_i, \mathbf{t})$
- (6) Calculate the probability of object \mathbf{obj}_i in the object set \mathbf{OBJ} , $\mathbf{P}(\mathbf{Visual}_m) = \mathbf{P}(\mathbf{obj}_n) = (\mathbf{S}(\mathbf{t}, \mathbf{t} + 1)_i) / \sum_{i=1}^n [\mathbf{S}(\mathbf{t}, \mathbf{t} + 1)_i]$.
- (7) **End**

ALGORITHM 1: VSIPAA algorithm.

experimental items, the probability of one or more experimental intents corresponding to them is also updated in real time.

4.3. Probability Model of Sensor Channel Based on Euclidean Distance. Distinct experimental actions correlate to different experimental intents in the sensor channel. During the experimental operation, the user dynamically generates a huge amount of complex sensor data, so this paper sets up a seven-dimensional vector γ to establish the mapping relationship among the three flex sensors, pressure sensors, and attitude sensors. On this basis, this paper designs the Euclidean distance sensor channel based intent probability acquisition algorithm (hereinafter referred to as SRIPAA) in the seven-dimensional space, as shown in Algorithm 2.

Under the sensor channel, different actions correspond to one or more experimental intents, and when the probability of the current user action is updated by the SRIPAA algorithm in this paper, the probability of the corresponding one or more experimental intents behind it is also updated simultaneously in real time.

4.4. Probability Model of Voice Channel Based on Task Slot. Users can enter speech information at any point during the experiment in the voice channel. This study divides speech_{*i*} into verb s_v and nouns s_n using Baidu voice recognition and

lexical analysis technology, and put them into the task slot to get the instruction V , as shown in Figure 5. Simultaneously, the voice command V performs similarity matching with the system speech library SPEECH task slot.

Under the voice channel, this paper designs a voice channel intention probability acquisition algorithm based on task slot (hereinafter referred to as VCIPPA), as shown in Algorithm 3.

4.5. Multimodal Fusion Algorithm. Multimodal fusion refers to the overall merging of all input information before the system confirms the user's intent in the context of a virtual-reality fusion experiment. The TMFA completes the experiment by applying the rule that one channel's input information corresponds to one experimental intent, which essence is that the serial integrates the diversity of single-mode information, rather than the parallel fusion of all modal intent at the intention level.

Therefore, this paper inventively proposes MFA based on the smart glove system, which calculates the coefficient of variation of each channel using the information weight method and obtains the weights of each channel by normalization, and then the joint intent probability of multimodal fusion is calculated and generated and achieves the function of fusing user multimodal intent information in parallel on the intent layer in the following steps:

Input: Fingertip pressure threshold ε , curvature value δ , pressure value vector group **Pressure**, curvature vector group **Curvature**, rotation angle θ and movement velocity \mathbf{v} generated by the posture sensor, action library **ACTION**.

Output: Experimental intent probability set **Sensor** under sensor channel.

- (1) **While Pressure! = Empty and Curvature! = Empty and θ and \mathbf{v} do.**
- (2) Vectorization of the pressure value information for each finger in **Pressure** = {**pressure_i**} ($1 \leq i \leq 5$), **if (Pressure_i $\geq \varepsilon$) \rightarrow pressure_i^T = [1, 1, 1, 1, 1] if (Pressure_i $< \varepsilon$) \rightarrow pressure_i^T = [0, 0, 0, 0, 0].**
- (3) Vectorization of the curvature value information for each finger in **Curvature** = {**curvature_i**} ($1 \leq i \leq 5$). The curvature dimension of the five-dimensional vector \mathbf{a} is divided into $[-180^\circ, -90^\circ)$, $[-90^\circ, -30^\circ)$, $[-30^\circ, 30^\circ)$, $[30^\circ, 90^\circ)$, $[90^\circ, 180^\circ)$, and when the curvature value δ_i of finger i is obtained, δ_i can be placed under the corresponding dimension of vector **curvature_i**.
curvature_i = Vectorization(δ_i).
- (4) Establishing the mapping between **Curvature** and **Pressure**, **pressure_i^T curvature_i = Mapping(Curvature, Pressure).**
- (5) Establishing the mapping of **Curvature**, **Pressure**, θ and \mathbf{v} , **$\gamma(\alpha_1^T \beta_1 \alpha_2^T \beta_2 \alpha_3^T \beta_3 \alpha_4^T \beta_4 \alpha_5^T \beta_5 \theta \mathbf{v}) = \text{Mapping(Curvature, Pressure, } \theta, \mathbf{v})$.**
- (6) Using the Euclidean distance formula in multidimensional space to find the distance between the vector γ and the action vector μ in the action library **ACTION**, **$\sqrt{\sum_{i=1}^7 (\gamma_i - \mu_i)^2} = \text{Distance E}_7(\gamma_i, \mu_i)$.**
- (7) Calculate the probability of the user performing action i in the action library **Action**, **$\mathbf{P}(\text{Sensor}_m) = \mathbf{P}(\text{action}_n) = (\text{Distance E}_7(\gamma_i, \mu_i) / \sum_{\text{ACTION}} \text{Distance E}_7(\gamma_i, \mu_i))$.**
- (8) **End**

ALGORITHM 2: SRIPAA algorithm.

Verb	Noun
------	------

FIGURE 5: Task slot.

Input: User-inputted voice **speech**, the system speech library **SPEECH**.

Output: Experimental intent probability set **Voice** under voice channel.

- (1) **While speech! = Empty do.**
- (2) Use the speech recognition and lexical analysis function to divide **speech** into verb \mathbf{s}_n and noun \mathbf{s}_v , $\mathbf{s}_n, \mathbf{s}_v = \text{participle}(\text{speech})$.
- (3) Fill \mathbf{s}_n and \mathbf{s}_v into the task slot to get the command **V**, **$\mathbf{V} = \text{slot}(\mathbf{s}_n, \mathbf{s}_v)$.**
- (4) Match the similarity between the instruction **V** and the system speech library **SPEECH**, and obtain the intent probability under the voice channel **$\mathbf{P}(\text{Voice}) = \begin{cases} 1, & \mathbf{V} = \text{SPEECH}, \\ 0, & \mathbf{V}! = \text{SPEECH}, \end{cases}$**
- (5) **End**

ALGORITHM 3: VCIPAA algorithm.

Input: Intent probability set **Sensor**, intent probability set **Voice**, intent probability set **Visual**.

Output: User's current intent **Intention**.

- (1) **While (Visual, Sensor, Voice)! = Empty or (Visual, Sensor)! = Empty do.**
- (2) Using the VSIPAA algorithm to obtain the intent probability set **Visual** under the visual channel, **Visual = VSIPAA(OBJ)**.
- (3) Using the SRIPAA algorithm to obtain the intent probability set **Sensor** under the sensor channel, **Sensor = SRIPAA(Pressure, Curvature, θ , \mathbf{v}).**
- (4) Using the VCIPAA algorithm to obtain the intent probability set **Voice** under the voice channel, **Voice = VCIPAA(speech)**
- (5) Calculate the probability mean of the set of probabilities of each channel intent, **$\mathbf{S}_{\text{Visual}}, \mathbf{S}_{\text{Sensor}}, \mathbf{S}_{\text{Voice}} = \text{Mean}(\text{Visual}, \text{Sensor}, \text{Voice})$.**
- (6) Calculate the probability variance of the set of probabilities of each channel intent, **$\mathbf{T}_{\text{Visual}}, \mathbf{T}_{\text{Sensor}}, \mathbf{T}_{\text{Voice}} = \text{Variance}(\text{Visual}, \text{Sensor}, \text{Voice})$.**
- (7) Using Algorithm 4 to calculate the coefficient of variation for each channel **D**, **$\mathbf{D}_i = \mathbf{S}_i / \mathbf{T}_i$.**
- (8) Normalize the coefficient of variation to obtain the weights of each channel, **$\omega_{\text{Visual}}, \omega_{\text{Sensor}}, \omega_{\text{Voice}} = \text{Normalization}(\mathbf{D}_{\text{Visual}}, \mathbf{D}_{\text{Sensor}}, \mathbf{D}_{\text{Voice}})$.**
- (9) Calculate the joint probability of the real-time intent of the three channels, and the final intent is the intent with the highest joint probability, as shown in Algorithm 4. **Intention = max{ $\omega_{\text{Visual}} \text{Visual}_i + \omega_{\text{Sensor}} \text{Sensor}_i + \omega_{\text{Voice}} \text{Voice}_i$ }.**
- (10) **End**

ALGORITHM 4: MFA.

4.6. Algorithm Analysis. Algorithm validity means that when the input value satisfies the condition, the algorithm should work properly and output the corresponding result; i.e., whether the breakpoint is set in the voice channel, the visual channel or the sensor channel will output the corresponding intent. In the real experiment, the user may not input the information of three channels at the same time; for example, the user may not input the voice during the operation. In that case, the algorithm will update the coefficient of variation of the visual and sensor channels in real time and flexibly to calculate the final intent probability.

When the user perceives the experimental scene information using the smart glove's camera on the wrist, the MFA employs YOLOv5 to identify the experimental objects in the scene and calculates the increase or decrease of the area of the experimental object between two frames in real time. This algorithm feeds the area increment change value of the experimental object's bounding to the VSIPAA algorithm, and the probability set Visual of the experimental intent is obtained under the visual channel.

When users use the smart glove to operate real or virtual experimental objects, the smart glove will continuously generate a large amount of unavailable and independent complex data in the process of operation, so the SRIPAA algorithm sets up a seven-dimensional vector γ to establish the mapping relationship between the three sensors, so the intent inference problem of the sensor channel can be converted into a problem of calculating the distance between vectors in the high-dimensional space. To acquire the experimental intent probability set Sensor under the sensor channel, the MFA utilizes SRIPAA algorithm to calculate the Euclidean distance of the vector in the high-dimensional space.

Under the voice channel, users can input different speech information speech at any time. And the MFA utilizes VCIPAA algorithm to generate input speech task slot V . By matching with the similarity of system speech library SPEECH, the intent probability set Voice under the voice channel can be easily calculated.

Throughout the experiment, the MFA changes the value of each channel's intent probability set in real time, as well as the corresponding weight value, i.e., the coefficient of variation, in response to the change in the probability value. Finally, based on the normalized coefficients of variation, the MFA performs an intention-level fusion of the probability sets of the three channels to produce the user's final intent.

The MFA efficiently handles two theoretically suggested problems in this paper: (1) the MFA provides an unobstructed real-time perception of scene information using the visual channel at the wrist of the smart glove compared to the indoor interactive input devices; (2) the MFA can fuse the user's multimodal intent in parallel at the intention level and improves on the TMFA serial processing of modal information.

5. Analysis of Experimental Results

5.1. Setting. The computer used to conduct the research in this paper had an i7-10875H processor with a 2.30 GHz

processor and 16 GB of RAM. The virtual-reality fusion lab was designed using Unity with version number 2018.3.8. To evaluate the effectiveness of the smart glove system with multimodal fusion in improving the teaching of MR experiments, 30 volunteers were invited to participate in the experiments while designing the comparison experiments in this paper. The volunteers included 10 elementary school students and 20 secondary school students with a male-to-female ratio of 1:1. The ages of these students were concentrated between 8 and 16 years old, and none of them had ever used the smart glove MR experiment system.

5.2. Experiment: Reduction of Iron Oxide by Charcoal. In this paper, a multimodal fusion of smart glove was utilized for the charcoal reduction of iron oxide experiment. And the system speech library of this experiment is shown in Figure 6. There are 11 user intentions in the whole experimental system: pick up the distilled water (*I1*), pour distilled water (*I2*), check air tightness (*I3*), pick up the clarified lime aqueous (*I4*), pour clarified lime aqueous (*I5*), pick up the spoon (*I6*), take out the charcoal powder (*I7*), add charcoal powder (*I8*), take out the iron oxide powder (*I9*), add iron oxide powder (*I10*), and turn on the alcohol burner (*I11*).

Meanwhile, according to the experimental requirements, this paper sets up the experimental intent library to analyze the user's experimental intent, as shown in Table 1.

The smart glove system receives the user's multimodal data in real time and converts it into the intent probability set of each channel throughout the user experiment operation. Next, Algorithm 4 performs a parallel fusion of intention levels on the intent probability set, before finally displaying the user's current experimental intent. As illustrated in Figure 7, this study establishes a multimodal output module [23] based on phenomenon visualization and speech output according to user intention information. After obtaining the user's intent, the module simulates the corresponding experimental phenomena, and guides and corrects the essential steps or incorrect steps to help the user through the experiment.

In this experiment, the second step is to pour distilled water. The user wears a smart glove, pours liquid from a real beaker into a virtual beaker, and inputs the voice "pour distilled water." Through Algorithm 4, the system deduces that the user's present behavior intends to "pour distilled water." The MR experiment system gives real-time feedback on the user's current behavior through the experiment scene animation, as shown in Figure 8(a). When the smart glove senses the user taking up the hot towel and covering the virtual beaker's wall, the system may still utilize Algorithm 4 to infer that the user's present behavior is "check air tightness" even without the voice input of "check air tightness." The experiment uses information augmentation technology to display the bubbles formed in the virtual beaker to demonstrate that the apparatus is well gasketed, as shown in Figure 8(b). As the operation proceeds, when the smart glove detects that the user picks up the medicine spoon to take chemicals from the fine mouth bottle containing iron oxide powder and enters the voice "take out iron

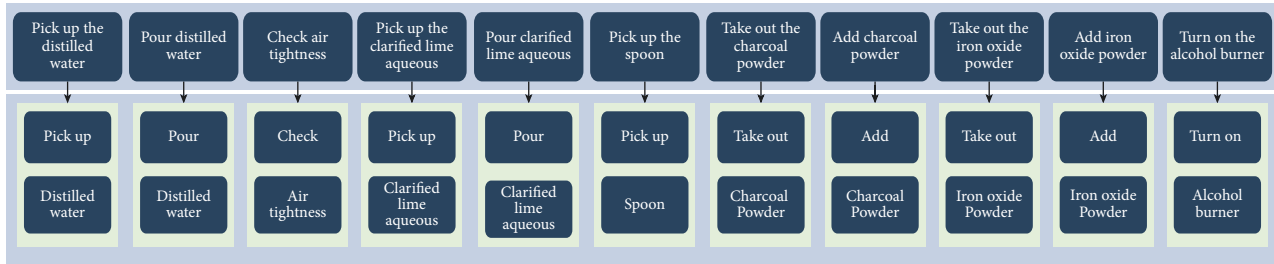


FIGURE 6: System speech library.

TABLE 1: Experiment library under voice channel, visual channel, and sensor channel.

Intent	System speech library task slot	Corresponding recognition object	Corresponding recognition action
I1	[Pick up, distilled water]	Beaker with water	Pick up
I2	[Pour, distilled water]	Beaker with water	Pour
I3	[Check, air tightness]	Hot towel	Grasp
I4	[Pick up, clarified lime aqueous]	Wide mouth bottle with clarified lime water	Pick up
I5	[Pour, clarified lime aqueous]	Wide mouth bottle with clarified lime water	Pour
I6	[Pick up, spoon]	Medicine spoon	Pinch
I7	[Take out, charcoal powder]	Fine mouth bottle with charcoal powder	Pinch
I8	[Add, charcoal powder]	Fine mouth bottle with charcoal powder	Pinch
I9	[Take out, iron oxide powder]	Fine mouth bottle with iron oxide powder	Pinch
I10	[Add, iron oxide powder]	Fine mouth bottle with iron oxide powder	Pinch
I11	[Turn on, alcohol burner]	—	Poke

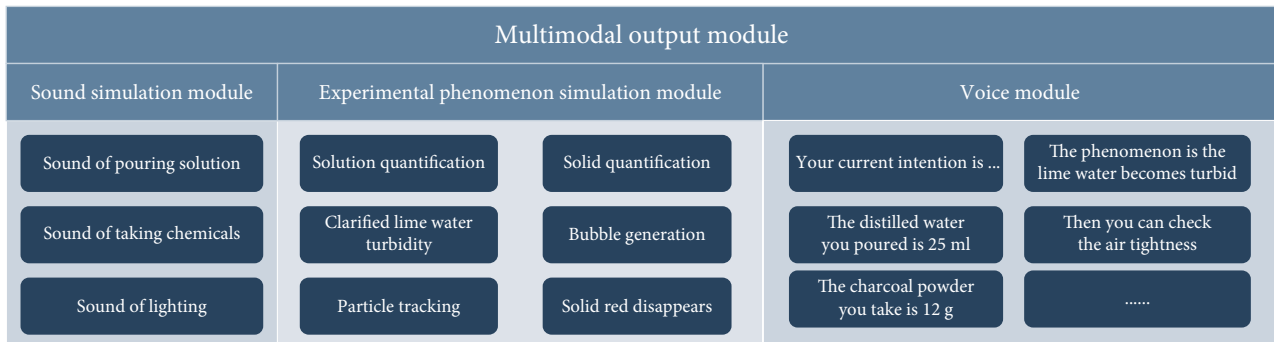


FIGURE 7: Multimodal output module.

oxide powder,” the smart glove system obtains the current operation behavior of the user as “take out the iron oxide powder.” The virtual iron oxide powder pellet in the experimental system follows the tip of the medication spoon, as shown in Figure 8(c). In the last step, after the user places the iron oxide powder and charcoal powder into the test tube, operates the virtual alcohol blowtorch, and at the same time enters the voice “light the alcohol blowtorch,” the system will get the current behavior of the user as “turn on the alcohol blowtorch.” As shown in Figure 8(d), the MR experiment system shows the virtual flame of the alcohol torch burning and the clarified lime water turning cloudy.

5.3. *Speech Instruction Verification.* The smart glove system can evaluate various speech information of users for voice

input. The VCIPAA algorithm translates the user’s speech information into instruction V and performs similarity matching with the system speech library, so as to update the intent probability set of the voice channel. Although users can input varied speech information into the smart glove system at any time, the intent probability of voice channel will change only when they input speeches similar to the system speech library. Consequently, to verify the stability of the voice channel, a comparative experiment is set up and 10 volunteers are invited to input the 11 instructions of the system speech library and irrelevant instructions, and the consequences of their recognition accuracy are tallied, as shown in Table 2.

Instead of wearing smart glove throughout the recognition process, the user can only input speech information, preventing information from other channels from interfering with the verification of voice channels. It should be

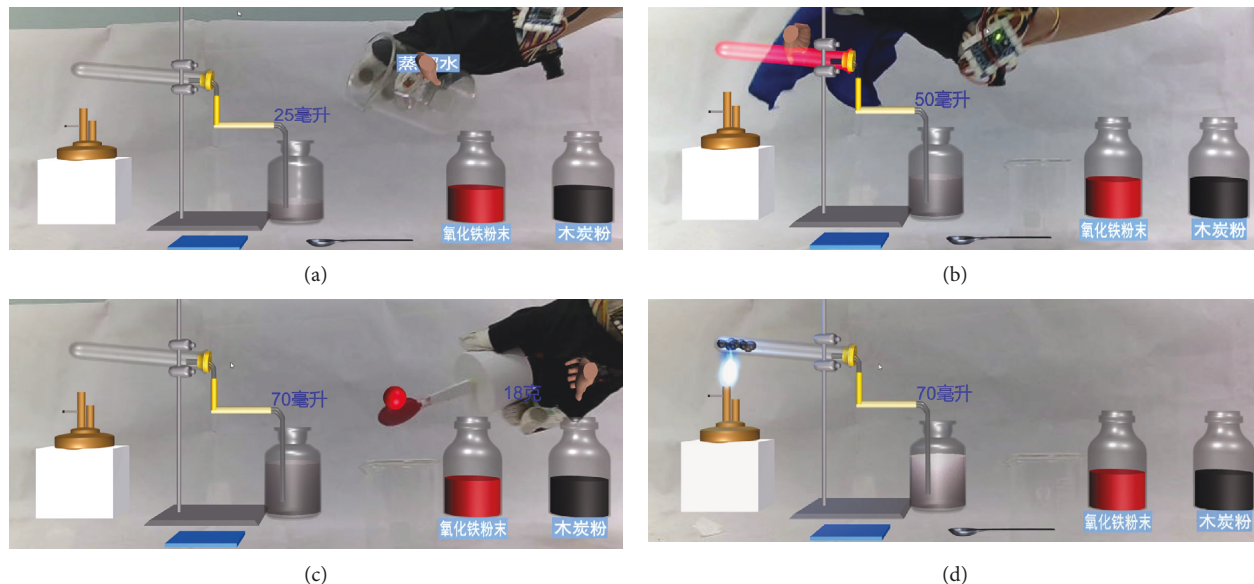


FIGURE 8: (a) A diagram of the user wearing a smart glove to pour distilled water. (b) A diagram of the user picking up a hot towel for gas tightness test. (c) A diagram of the user removing the iron oxide powder after the grains follow the movement of the tip of the spoon. (d) A diagram of the experimental phenomenon after the user ignites the alcoholic blowtorch.

TABLE 2: Relevant instructions and irrelevant instructions.

	Relevant instructions	Irrelevant instructions
Total times	110	110
Effective times	107	0
Recognition accuracy	97.27%	0%

highlighted that a successful recognition instruction indicates that the MR experimental system will display the related experimental phenomenon when the user inputs the corresponding speech.

As shown in Table 2, the MR system will not respond if the user enters irrelevant instructions. However, the recognition accuracy reaches 97.27% when users input the relevant instructions of the system speech library, but it is also affected by network delay or nonstandard user pronunciation. To summarize, the speech channel has a high level of stability, which is one of the foundations for MFA accurately identifying the user's experimental intent.

5.4. Occlusion Handling. One of the biggest issues with traditional indoor interactive input devices in the context of intelligent experiments is occlusion. The instruments and devices in the experimental scenario are often occluded due to different placements, making the input device unable to perceive and recognize key information, especially difficult to recognize tiny objects and observe subtle phenomena. Traditional virtual-reality fusion chemistry experiments [2] used KINECT for user gesture channel information acquisition and a binocular camera

for visual channel information acquisition. As a result, it is easy to have a problem during the experiment with improper recognition or occlusion of experimental objects and gestures, which impacts the impression of the user's intent, as shown in Figure 9. Therefore, this paper uses the camera at the wrist of the smart glove for proximity perception of the scene to solve the problems existing in the traditional virtual laboratory.

In this paper, 30 volunteers were invited to participate in a comparative experiment in which they completed several key steps of both the traditional virtual-reality fusion experiment and the MR experiment based on smart glove system. Each volunteer conducted 3 times, their recognition accuracy was counted (the recognition success here means that the virtual-reality fusion platform can show the correct experimental phenomenon), and the results are shown in Figure 10.

As can be learned from Figure 10, compared with the traditional virtual-reality fusion experiment, the smart glove system has a significant advantage in terms of observation of tiny objects, subtle movements, and minute phenomena.

For the action of picking up a medicine spoon, it is difficult for KINECT to respond to it because the spoon is small and not easy to recognize, but the wrist camera of smart glove system can easily recognize the object in the experimental scene. For the action of turning on the alcohol burner, the magnitude of the action is small, so the KINECT device can easily recognize the gesture as other actions. However, the smart glove system can employ MFA to fuse multichannel information to recognize the action.

In summary, the smart glove system can solve the problem that it is difficult to cope with subtle actions and micro-phenomena in the traditional virtual-reality fusion experiment.

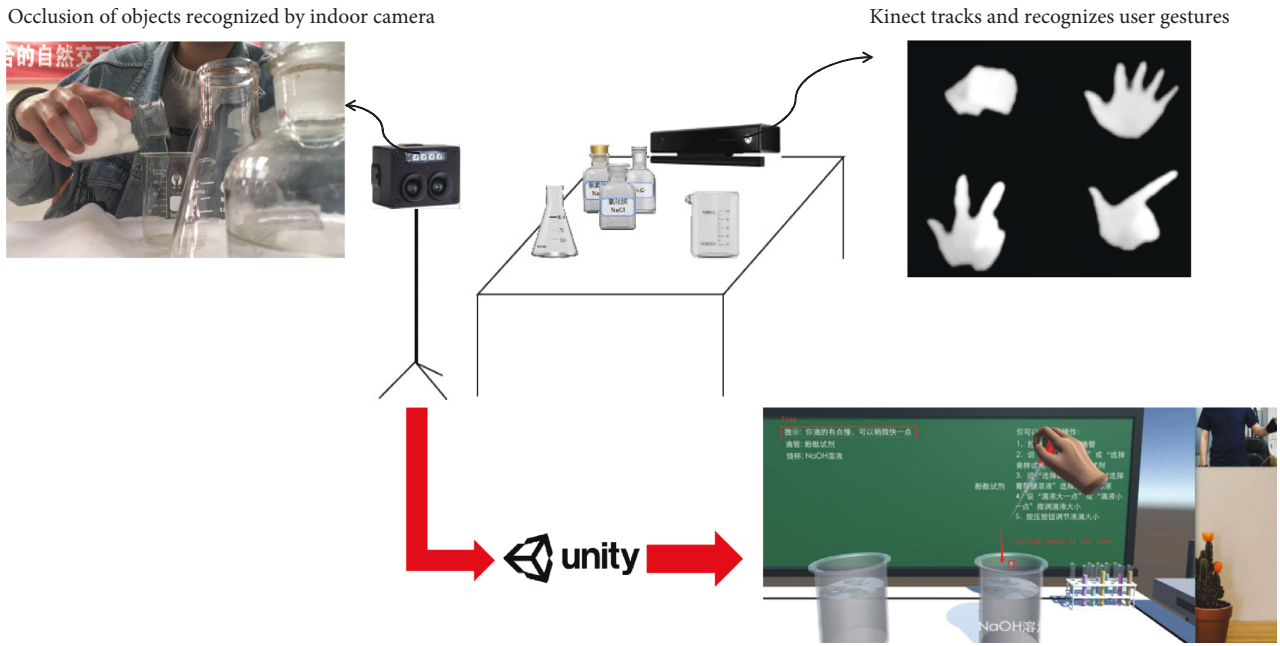


FIGURE 9: Diagram of the occlusion problem.

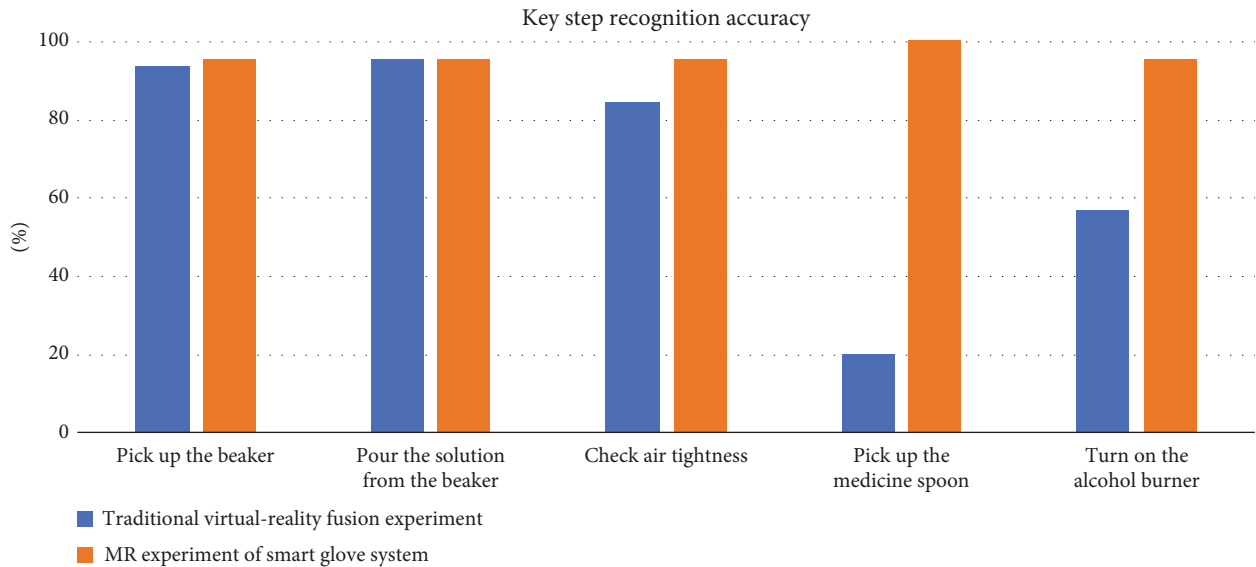


FIGURE 10: Key step recognition accuracy.

5.5. *Verification of Multimodal Fusion.* Users can complete the MR chemical experiment by using a single-mode interaction with an independent voice channel or by combining two or three modes. However, during the experiment, the visual and sensor channels will continuously collect data from the user, while the voice channel will be input according to the user's choices, allowing the user to choose the number of channels to employ based on the current experimental steps.

In order to verify that multimodal fusion can better identify user experimental intent than single mode, 20

volunteers were invited to complete the experiment 4 times using (1) voice channel, (2) dual-mode fusion of visual channel and sensor channel, and (3) multimodal fusion of three channels, and counted the relationship between the average completion rate, average completion time (in seconds), and user satisfaction (10-point scale) to verify that multimodal fusion can better identify users' intentions than single-mode fusion. (The completion experiment here means that the MR platform displays the operation's reaction outcomes in real time.) Table 3 displays the validation findings.

Table 3 can be used to draw the following conclusions:

TABLE 3: Comparison of three methods.

Intent name	Voice channel	Dual-mode fusion	Multimodal fusion
Average completion rate	97.5%	88.75%	95%
Average completion time	71	274	201
User satisfaction	3.9	6.05	8.9

- (1) When the user uses the voice channel alone, the experiment completion rate is 97.5% since the input voice information is rather constant. Users, on the other hand, are unable to operate genuine experimental objects, which result in the user experience is terrible.
- (2) Because an experimental object in the visual channel and an action in the sensor channel might both correspond to separate experimental intentions at the same time, the average completion rate drops, but the user's score improves significantly when compared to the single voice channel. Because the data from the sensor and visual channels are complicated, the temporal complexity of dual-mode fusion increases, resulting in a longer average completion time for users.
- (3) Due to the nonstandard operation of the user in the multimodal fusion process, the average completion rate of the experiment will be lower than that of utilizing only the voice channel, but the user score will be the highest in the comparative experiment. Furthermore, when compared to dual-mode fusion, after the addition of a voice channel, the average completion time of a multimodal fusion experiment is lowered. So, it is shown that the experimental interaction mode of multimodal fusion can suit users' needs and can better comprehend their experimental intentions.

5.6. Verification of MFA Algorithm. Although the TMFA [11] uses multichannel data information in the experimental process, its essence is serial fusion of multimodal information, which means that only one channel of information is used in each intent recognition. The essence of MFA is the parallel fusion of multimodal intent probability. In order to verify if the MFA has a superior intent recognition impact than TMFA, 15 volunteers were invited to complete experiments of iron oxide reduction with charcoal 5 times using the two algorithms in a comparison study. Figure 11 shows the average intention recognition rate and the cost time (in seconds) of the experiment.

Figure 11 shows that the average intent recognition rate of the two algorithms in each experiment is more than 90%, demonstrating that the two algorithms can accurately identify the user's intention and that the user's cost time to complete the experiment with the help of the two algorithms decreases as the experiment progresses. However, when compared to the TMFA, the MFA has a higher average intent recognition rate and reduces the time to complete the experiment by around one minute, demonstrating that the

MFA can assist users in accurately and efficiently completing the experiment, and can be applied to middle school chemistry experiment education.

5.7. Cognitive Burden Assessment. To highlight the ability of the smart glove to perceive the behavioral intent of users, a comparison experiment was designed in this paper. First, volunteers conducted experiments on a virtual experimental platform using the Noitom data glove [24] and the KINECT device; second, volunteers perform the same experiment on the NOBOOK platform [1], which is dominated by keyboard and mouse operations; third, volunteers use the traditional virtual-reality fusion experimental platform [2] to conduct experiments; at last, volunteers perform the same experiments using the smart glove. 30 volunteers were required to take turns in conducting experiments on the above four experimental platforms during a day and to perform NASA evaluation [25] of each experiment after completion. User evaluation metrics were categorized into mental demands (MD), physical demands (PD), time demands (TD), performance (P), effort (E), and frustration (F). The NASA assessment metrics were evaluated on a 5-point scale. 0 to 1 indicates a low cognitive burden, 1 to 2 indicates a relatively low cognitive burden, 2 to 3 indicates an overall cognitive burden, 3 to 4 indicates a relatively high cognitive burden, and 4 to 5 indicates a very high cognitive burden. The results are shown in Figure 12.

Figure 12 shows that the MD and TD index scores of smart glove are lower than those of other platforms, indicating that the experimental process of smart glove is simpler than that of others. This is because when using other platforms for experiments, volunteers need to understand the various functions of the platform in advance, such as the construction of the NOBOOK experimental platform. The smart glove system and the virtual-reality fusion experimental platform scored higher on the P index. When it came to running the experiment, volunteers indicated they spent the majority of their time learning how to utilize other two platforms. When they wear smart glove for the experiment, they pay more attention to the phenomenon and results of the experiment. The experimenter can deepen their understanding of the experimental phenomenon by observing the phenomenon on the screen and the system's explanation of the experimental mechanism. At the same time, the smart glove system will also correct the nonstandard behavior in the experimental process, so that they will have a deeper impression of the key points of the experimental operation.

In conclusion, as compared to existing experimental platforms, the smart glove system developed in this paper

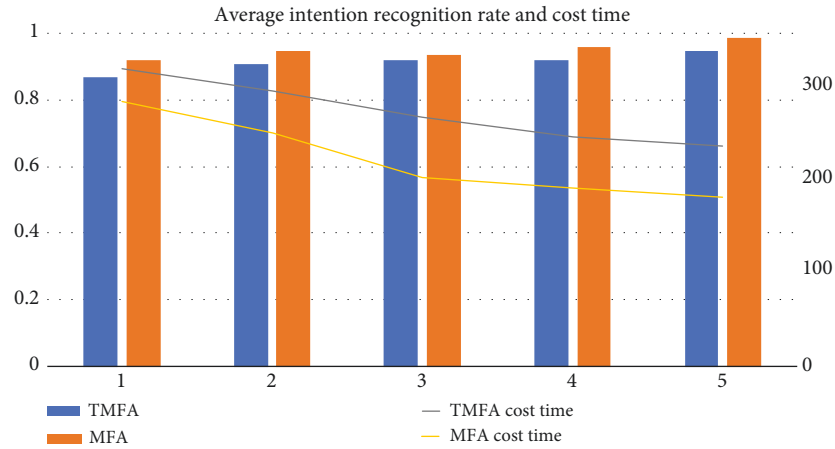


FIGURE 11: Comparison between MFA and TMFA.

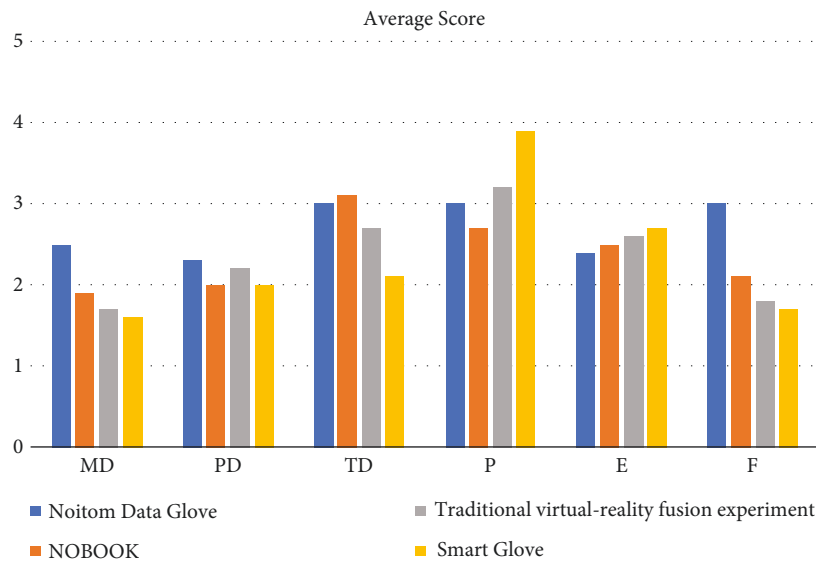


FIGURE 12: NASA user evaluation.

allows users to conduct experiments more intelligently and naturally, as well as improves users’ experimental immersion and operation ability.

6. Conclusions

Some middle school experiments currently have issues with powerful destruction and expensive costs, while the traditional experimental platform suffers from a poor sense of operation and a significant memory burden. As a result, this paper develops a smart glove using the MFA to detect the user’s experimental intent and then direct the user to experiment, or correct and remind the user’s incorrect and harmful actions.

To address the aforementioned issues, this study primarily performs the following two original contributions:

- (1) This study designs and implements a smart glove system. The smart glove can address the issues of (a) the lack of cognitive ability of traditional data gloves and (b) the occlusion phenomenon caused by too many indoor interactive input devices in the traditional virtual-reality fusion chemistry experiment. The traditional data glove can only gather the user’s hand data and cannot capture the user’s experimental intent; thus, several indoor interactive input devices are required to aid the user during the perception process of scene information in real time, which can easily cause occlusion in object recognition. However, the smart glove system can realize the observation of tiny objects, subtle movements, as well as minute phenomena, increase interaction efficiency, and decrease memory pressure. However,

the smart glove's hardware arrangement is unreasonable, and it is difficult to wear when the user's palm is large; the MR system is currently just for chemical research. As a result, future research should focus on improving the smart glove's structure and expanding the experiment library so that users can complete a wide range of experiments in order to achieve the goal of virtual-reality education [26].

- (2) A parallel MFA integrating sensor, voice, and visual channels is proposed in this paper to address the problem that the TMFA only integrates user modal information in serial. The essence of TMFA is the serial fusion of users' multimodal information, and it does not support the simultaneous analysis of users' intentions using multimodal data. However, the MFA is a multimodal fusion algorithm based on multichannel intention probability fusion, which can accurately and efficiently obtain the user's intent and guide user's behavior based on that intent. Based on the user's intent, future research should be devoted to predicting the user's intent in the following stage and building a human-computer cooperation model so that the user can wear the smart glove and cooperate with the computer to complete experiments.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Independent Innovation Team Project of Jinan City (no. 2019GXRC013) and the Natural Science Foundation of Shandong Province (no. ZR2020LZH004).

References

- [1] S. Fang, L. Xue, Y. Liang, J. Wu, C. Ju, and M. Zhao, "NOBOOK VR experiment test," in *Proceedings of the 2020 international conference on virtual reality and visualization (ICVRV)*, pp. 346-347, IEEE, Recife, Brazil, November 2020.
- [2] B. Zeng, Z. Feng, T. Xu, and R. Han, "Research on intelligent experimental equipment and key algorithms based on multimodal fusion perception," *IEEE Access*, vol. 8, Article ID 142507, 2020.
- [3] K. Aljuhani, M. Sonbul, M. Althabiti, and M. Meccawy, "Creating a Virtual Science Lab (VSL): the adoption of virtual labs in Saudi schools," *Smart Learning Environments*, vol. 5, no. 1, pp. 1-13, 2018.
- [4] D. Bogusevski, C. Muntean, and G. M. Muntean, "Teaching and learning physics using 3D virtual learning environment: a case study of combined virtual reality and virtual laboratory in secondary school," *Journal of Computers in Mathematics and Science Teaching*, vol. 39, no. 1, pp. 5-18, 2020.
- [5] P. Salinas and R. Pulido, "Visualization of conics through augmented reality," *Procedia Computer Science*, vol. 75, pp. 147-150, 2015.
- [6] L. Lenz, D. Janssen, and V. Stehling, "Mixed Reality Voice Training for lecturers," in *Proceedings of the 2017 4th Experiment@ International Conference (Exp. At'17)*, pp. 107-108, IEEE, Faro, Portugal, June 2017.
- [7] Z. Hu, Y. Zhang, Y. Xing, D. Cao, and C. Lv, "Toward Human-Centered Automated Driving: A Novel Spatiotemporal Vision Transformer-Enabled Head Tracker," *IEEE Vehicular Technology Magazine*, 2022.
- [8] S. A. Chhabria, R. V. Dharaskar, and V. M. Thakare, "Survey of Fusion Techniques for Design of Efficient Multimodal systems," in *Proceedings of the 2013 International Conference on Machine Intelligence and Research Advancement*, pp. 486-492, IEEE, Katra, India, December 2013.
- [9] A. W. Ismail, M. Billingham, M. S. Sunar, and C. S. Yusuf, "Designing an Augmented Reality Multimodal Interface for 6DOF Manipulation techniques," *SAI Intelligent Systems Conference*, Springer Cham, Switzerland, England, 2018.
- [10] M. S. Kadavasal and J. H. Oliver, "Virtual reality interface design for multi-modal teleoperation," in *Proceedings of the ASME world conference on innovative virtual reality*, pp. 169-174, Article ID 43376, Chalon-sur-Saône, France, January 2009.
- [11] M. Xiao, Z. Feng, X. Fan, B. Zeng, and J. Li, "A structure design of virtual and real fusion intelligent equipment and multimodal navigational interaction algorithm," *IEEE Access*, vol. 8, Article ID 125982, 2020.
- [12] H. Liu, T. Fang, T. Zhou, and L. Wang, "Towards robust human-robot collaborative manufacturing: multimodal fusion," *IEEE Access*, vol. 6, Article ID 74762, 2018.
- [13] Z. Hu, Y. Xing, W. Gu, D. Cao, and C. Lv, "Driver Anomaly Quantification for Intelligent Vehicles: A Contrastive Learning Approach with Representation Clustering," *IEEE Transactions on Intelligent Vehicles*, 2022.
- [14] S. Ondas, J. Juhar, M. Pleva, and R. Holcer, "Service robot SCORPIO with robust speech interface," *International Journal of Advanced Robotic Systems*, vol. 10, no. 1, p. 3, 2013.
- [15] P. Lokhande, R. Prajapati, and S. Pansare, "Data gloves for sign language recognition system," *International Journal of Computer Application*, vol. 975, p. 8887, 2015.
- [16] K. Abhijith Bhaskaran AnoopK. Deepak Ram, A. Krishnan, and H. R. Nandi Vardhan, "Smart gloves for hand gesture recognition: sign language to speech conversion system," in *Proceedings of the 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA)*, pp. 1-6, IEEE, Amritapuri, India, December 2016.
- [17] C. Kumar Mummadi, F. P. P. Leo, K. Deep Verma, S. Kasireddy, P. Marcel Scholl, and K. Van Laerhoven, "Real-time embedded recognition of sign language alphabet fingerspelling in an imu-based glove," in *Proceedings of the 4th International Workshop on Sensor-Based Activity Recognition and Interaction*, pp. 1-6, ACM, Broadway, NY, USA, September 2017.
- [18] K. Roy, D. Prasad Idiwali, A. Agrawal, and B. Hazra, "Flex sensor based wearable gloves for robotic gripper control," in *Proceedings of the 2015 Conference on Advances in Robotics (AIR '15)* New York, NY, USA, ACM, July 2015.
- [19] Z. Ma, P. Ben-Tzvi, and J. Danoff, "Hand rehabilitation learning system with an exoskeleton robotic glove," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 12, pp. 1323-1332, 2015.

- [20] Q. Liu, M. Q. Guo, W. Meng, Q. Ai, C. Yin, and Z. Fang, "A new IMMU-based data glove for hand motion capture with optimized sensor layout," *International Journal of Intelligent Robotics and Applications*, vol. 3, no. 1, pp. 19–32, 2019.
- [21] Y. Ge, B. Li, W. Yan, and Y. Zhao, "A real-time gesture prediction system using neural networks and multimodal fusion based on data glove," in *Proceedings of the 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*, pp. 625–630, Xiamen, China, March 2018.
- [22] A. Bandura, "Social cognitive theory of self-regulation," *Organizational Behavior and Human Decision Processes*, vol. 50, no. 2, pp. 248–287, 1991.
- [23] S. Ondáš, J. Juhár, M. Pleva, P. Ferčák, and R. Husovský, "Multimodal dialogue system with NAO and VoiceXML dialogue manager," in *Proceedings of the 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, IEEE, Debrecen, Hungary, September 2017.
- [24] P. Neuron, "Perception Neuron by Noitom| Perception Neuron Motion Capture for Virtual Reality, Animation, Sports, Gaming and film," 2018, <https://neuronmocap.com/>.
- [25] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Advances in Psychology*, vol. 52, pp. 139–183, 1988.
- [26] M. Pleva, J. Juhar, S. Ondas, C. R. Hudson, C. L. Bethel, and D. W. Carruth, "Novice User Experiences with a Voice-Enabled Human-Robot Interaction tool," in *Proceedings of the 2019 29th International Conference Radioelektronika (RADIOELEKTRONIKA)*, pp. 1–5, IEEE, Pardubice, Czech Republic, April 2019.

Research Article

Deep Multi-Scale Residual Connected Neural Network Model for Intelligent Athlete Balance Control Ability Evaluation

Nannan Xu,¹ Xin Wang ,² Yangming Xu,³ Tianyu Zhao ,⁴ and Xiang Li ⁵

¹Sports Training Institute, Shenyang Sport University, Shenyang 110115, China

²Department of Kinesiology, Shenyang Sport University, Shenyang 110115, China

³School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China

⁴Key Laboratory of Structural Dynamics of Liaoning Province, College of Sciences, Northeastern University, Shenyang 110819, China

⁵Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an 710049, China

Correspondence should be addressed to Xin Wang; wangxin@syty.edu.cn and Tianyu Zhao; zhaotianyu@mail.neu.edu.cn

Received 29 March 2022; Accepted 18 April 2022; Published 26 May 2022

Academic Editor: Jie Liu

Copyright © 2022 Nannan Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Athlete balance control ability plays an important role in different types of sports. Accurate and efficient evaluations of the balance control abilities can significantly improve the athlete management performance. With the rapid development of the athlete training field, intelligent and automatic evaluations have been highly demanded in the past years. This study proposes a deep learning-based athlete balance control ability evaluation method through processing the time-series movement pressure measurement data. An end-to-end model structure is proposed, which directly analyzes the raw data and provides the evaluation results, which largely facilitates practical utilization. A multi-scale feature extraction scheme is employed, by exploring the learned features in different scales. A residual connected neural network architecture is further proposed. By using the short-cut connection, the deep neural network model can be more efficiently trained. Experiments on the real athlete balance control ability tests are carried out for validations. Through comparisons with different related methods, the results show the proposed deep multi-scale residual connected neural network model is well suited for the athlete balance control ability evaluation problem, and promising for actual applications in the real scenarios.

1. Introduction

Balance control ability is of great importance for athletes. A number of sport areas with precise movement require accurate and efficient evaluations of the balance control ability for the athletes, such as freestyle skiing aericals, skating, and so forth [1]. Good evaluations of the balance control ability can well support the management of the athletes, including selection, training, competition, and so on. Accurate evaluation of the balance control ability remains a quite challenging issue, since a large number of factors are included, and the underlying ability cannot be well reflected. Significant expert knowledge and human labor are also highly required for this task, which makes it difficult to be carried out in the practical scenarios [2].

In the recent years, with the rapid development of the sensing technologies and data analysis methods, data-driven athlete balance control ability evaluation becomes feasible [3]. Specifically, the movement pressure measurement machine, such as a balance meter, can be used to collect the athlete subtle movement when they are standing on the machine. The collected signal can be used to evaluate the athlete balance control ability, since smaller movement pressure generally indicates better balance control ability, while larger movement pressure means the balance control ability is at lower level [4].

With respect to the collected data, typical statistical features can be used for balance control ability evaluation, such as mean, root mean square, and so forth [5–7]. In the past years, many signal processing methods have

been proposed for better feature extraction [8–10], including wavelet analysis, stochastic resonance techniques, and so on. Some machine learning and statistical inference techniques are also developed for solving the pattern recognition problem, such as artificial neural networks (ANN) [11], support vector machines (SVM), random forest, fuzzy inference, and so on [12–14]. However, the collected movement pressure data usually contain much noise, which makes it difficult to use the conventional features for evaluations. Furthermore, for the high-level athletes, the difference between different levels of athlete on balance control ability is quite small. The typical features cannot well reflect the difference. Therefore, the traditional data-driven methods on athlete balance control ability evaluation are facing great challenges at present.

Deep neural network has been the emerging technologies of artificial intelligence in the past years [15–23], and it has achieved great successes in many applications such as image recognition [24–26] and speech recognition [27]. Driven by big data, deep neural network can well learn the mapping function between the input data and the output pattern automatically [28–30]. High prediction accuracy can be usually obtained. In addition, deep neural network is generally a black box tool for automatic computations, which requires little prior knowledge on signal process or domain expertise [31]. Therefore, it is quite promising for solving the challenging athlete balance control ability evaluation problem.

The recent studies [32–35] show the time-series data can be well processed by the deep neural network model, and higher feature extraction efficiency and better effects can be generally obtained using deep learning [36, 37]. Different types of time-series data have been successfully processed using deep learning, including the medical data, financial data, condition monitoring data, and so on. Miao [38] proposed a deep learning framework for continuous blood pressure measurement using one-channel ECG signal. Promising effects have been obtained for processing the pressure data. An end-to-end intelligent morphological classification method for intracranial pressure pulse waveforms was proposed in the studies [39], where the deep learning method was applied for automatic feature extraction and pattern learning.

However, the typical deep neural network model suffers from many factors. For instance, the training efficiency is generally weak with the deep architecture [10, 40–45]. Traditional model establishment approach basically loses feature information in the feed-forward manner with a single-scale feature extraction scheme. The limitations hinder the development of the deep neural network methods.

In this study, a novel deep multi-scale residual connected neural network model is proposed to address the athlete balance control ability evaluation problem, as well as the remaining problems of deep neural networks. The main novelties and contributions of this study are listed as follows:

- (i) A new multi-scale feature extraction scheme is proposed, which consists of automatic feature learning in different scales. The integration of multi-scale features further enhances the information fusion performance and leads to better results.
- (ii) A deep residual connected module is proposed, which introduces short-cut connection between different convolutional layers in the deep neural network model. In this way, the training efficiency can be largely enhanced.
- (iii) The athlete balance control ability evaluation problem is investigated, and an intelligent method is proposed to achieve automatic feature extraction and evaluation. This has been seldomly studied in the current literature, and this study provides new insight in this task.
- (iv) Experiments on the real athlete under-feet movement pressure measurement data are used for validations of the proposed method. The results show that the proposed method can achieve high evaluation accuracy, and promising for applications in the real scenarios.

This study starts with the description of the preliminaries in Section 2. The proposed deep multi-scale residual connected method is presented in Section 3. Experiments are carried out for validations of the proposed method, and the results are shown in Section 4. We close the study with conclusions in Section 5.

2. Preliminaries

In this section, the preliminaries that are used in this study are presented, including the convolutional neural network, pooling, and softmax function. The concerned problem in this study can be formulated as learning a mapping function, which projects the raw collected athlete time-series data to the corresponding balance control ability level. The relationship is complex, and the traditional methods cannot well address this problem. Therefore, we propose a deep learning-based approach for modeling the highly nonlinear relationship.

2.1. Convolutional Neural Network. Convolutional neural networks (CNNs) have been one of the most popular neural network structures in the current literature. The effectiveness of CNNs has been widely validated in many application scenarios, such as the image classification tasks, speech recognition problems, and video processing tasks [46]. The variable and complicated signals can be automatically processed using CNNs, and high-level features can be effectively extracted. In the recent years, many researches have been carried out using CNNs and achieved significant successes [43, 47].

The most representative features of CNNs are the local receptive fields and shared parameters in signal processing. The data shift of the input data can be efficiently filtered out during feature extraction, and the spatial sub-sampling

algorithm can well extract the most remarkable features from the collected data. In this study, CNNs are used as the main framework for the data-driven intelligent feature extraction of the signal.

To be specific, the convolutional layers are placed to convolve different filters with respect to the raw data, and high-level features can be obtained accordingly. In most cases, the pooling operations are used after the convolutional operations, which can further extract the most significant features for the following processing. Meanwhile, the feature dimension can be also well reduced, which benefits the processing costs.

In this study, the data are a sequence of the time-series collections. Therefore, the 1-dimensional (1D) CNN is mostly adopted for the data processing, and that will be presented in the following. Let $\mathbf{x} = [x_1, x_2, \dots, x_N]$ denote the input data, where N represents the dimension of the input data sample. The convolutional computation can be defined using the filter kernel \mathbf{w} , $\mathbf{w} \in R^{F_L}$, where F_L represents the size of the filter kernel, defining the dimension of the local receptive field. The concatenation vector $\mathbf{x}_{i:i+F_L-1}$ can be defined as

$$\mathbf{x}_{i:i+F_L-1} = x_i \oplus x_{i+1} \oplus \dots \oplus x_{i+F_L-1}, \quad (1)$$

where the item $x_{i:i+F_L-1}$ is defined as the window with F_L sequential data points starting from the i -th data point. The operation \oplus is for concatenating the concerned data into a larger information entity. At last, the convolution computation can be expressed as

$$k_i = \eta(\mathbf{w}^T \mathbf{x}_{i:i+F_L-1} + m). \quad (2)$$

In this equation, m and η denote the bias vector and the neural network activation function, respectively. The feature map output k_i is known as the obtained features with respect to the filter kernel. Through applying the filter kernel from the first data point to the end on the input data sample, the learned feature representation can be calculated as

$$\mathbf{k}_j = [k_j^1, k_j^2, \dots, k_j^{N-F_L+1}]. \quad (3)$$

The expressions above represent the learned features. In the actual applications of the CNNs, a number of convolutional kernels can be used in one layer to obtain richer information from the raw data.

2.2. Pooling. In the typical neural networks, after the convolutional layer, a pooling layer is usually used for further feature extraction with respect to the learned feature maps. There are mainly two reasons for the utilization of pooling operations. First, the most significant features can be usually extracted by using the simple pooling functions, which provides an easy way for efficient learning. Second, the dimension of the feature maps can be largely reduced, which can help increase the processing efficiency. In this study, the max-pooling function is used, which has been popularly adopted in the literature for the related classification problems. Let p denote the size of the pooling operation. With respect to the extracted feature maps from

the convolutional layers, the pooled features can be expressed as

$$\begin{aligned} \mathbf{q}_j &= [q_j^1, q_j^2, \dots, q_j^s], \\ q_j^z &= \max(k_j^{(z-1)p+1}, k_j^{(z-1)p+2}, \dots, k_j^{zp}), \end{aligned} \quad (4)$$

where \mathbf{q}_j represents the obtained features from the pooling operation on the j -th feature map that has the size of s .

2.3. Softmax Function. Softmax function is a popular function in the data-driven neural network-based classification tasks. It is usually adopted at the end layer of the deep neural network. The values of the neurons can be transformed to the predicting probabilities by using the softmax function [25]. Specifically, after multiple combinations of convolutional and pooling layers in the deep neural network, the final extracted features are the input of the softmax function. Let $\mathbf{x}^{(i)}$ denote the training samples, and $r^{(i)}$ denote the corresponding class labels of the training samples. $i = 1, 2, \dots, N_{tr}$, where N_{tr} represents the training data sample number. We also have $\mathbf{x}^{(i)} \in R^{N \times 1}$ and $r^{(i)} \in \{1, 2, \dots, B\}$, where B represents the total number of concerned classes in the problem. With respect to the input data sample $\mathbf{x}^{(i)}$, the softmax function can well predict the class probability $p(r^{(i)} = j | \mathbf{x}^{(i)})$, $j = 1, 2, \dots, B$ for different class labels. The calculated probabilities of the data samples for each class can be computed based on the hypothesis function

$$\begin{aligned} J_\lambda(\mathbf{x}^{(i)}) &= \begin{bmatrix} p(r^{(i)} = 1 | \mathbf{x}^{(i)}; \lambda) \\ p(r^{(i)} = 2 | \mathbf{x}^{(i)}; \lambda) \\ \vdots \\ p(r^{(i)} = B | \mathbf{x}^{(i)}; \lambda) \end{bmatrix}, \\ &= \frac{1}{\sum_{b=1}^B e^{\lambda_b^T \mathbf{x}^{(i)}}} \begin{bmatrix} e^{\lambda_1^T \mathbf{x}^{(i)}} \\ e^{\lambda_2^T \mathbf{x}^{(i)}} \\ \vdots \\ e^{\lambda_B^T \mathbf{x}^{(i)}} \end{bmatrix}, \end{aligned} \quad (5)$$

where $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_B]^T$ represents the function coefficients. It can be noted that the softmax function classifier guarantees that the output values are all positive and the sum of them is one. Therefore, the softmax function is able to transform the outputs of the deep neural network to be the predicted probabilities for different concerned classes.

3. Proposed Deep Multi-Scale Residual Connected Model

In this study, a novel deep learning-based multi-scale residual connected model is proposed for time-series data processing and athlete performance evaluation. In this

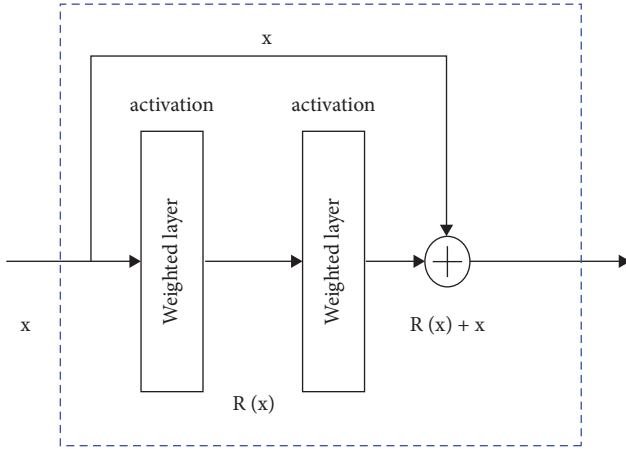


FIGURE 1: The proposed residual connected scheme in the deep neural network framework.

section, the proposed method is illustrated in detail, which consists of residual connection, multi-scale feature extraction, and end-to-end relationship model.

3.1. Residual Connection. In the traditional deep neural network, the back-propagation optimization method is usually used for model parameter updates. However, as the model architecture is typically deep with multiple layers, the optimization efficiency is not satisfactory in most cases due to the gradient vanishing problem, which makes the deep neural network difficult to achieve the optimal performance. Therefore in this study, a residual connected neural network scheme is proposed, which is illustrated in Figure 1. The residual connected module generally consists of three main characteristics.

- (i) A short-cut connection is used, which makes the information of the data can propagate through different layers, and directly into the subsequent layers in the network.
- (ii) With the residual connected module, deep neural network architecture can be adopted, since the gradient vanishing problem can be largely solved.
- (iii) The residual connected module is a relatively independent module with respect to the deep neural network structure, which can be readily added and removed from the existing architecture. Limited additional costs will be introduced for using the residual connected module.

Specifically, the residual connected module can be defined as

$$\mathbf{c} = R(\mathbf{x}, \{\mathbf{v}_i\}) + \mathbf{x}, \quad (6)$$

where \mathbf{x} and \mathbf{c} denote the input data and output data for the layer, respectively. The function R represents the residual connected operation. For example, $R = \mathbf{v}_2 \eta(\mathbf{v}_1^T \mathbf{x})$ can be used for a simple structure with the weights \mathbf{v}_i . The practical implementation of the residual connected operation is realized by the short-cut and element-wise sum. The non-

linear activation function can be used either before the sum or after the sum.

3.2. Multi-Scale Feature Extraction. In this study, a multi-scale feature extraction scheme is proposed to better learn the new features from the raw collected data. Specifically, the filter size in the convolutional operation plays an important role in the automatic feature learning process. Large filter size indicates that the learned features are more general and global with respect to the input data. Correspondingly, smaller filter size means the model pays more attention on the local features. In the current literature, there is no general consensus of the optimal selection of the filter size. Therefore, in this study, we propose to use multiple filter size for the feature extraction, in order to both take advantage of the global and local features from the input data.

In the deep neural network structure, three data and feature streaming approaches are proposed as shown in Figure 2. In each approach, a certain size of the convolutional filter is utilized. The common range of the filter sizes is covered in this study, and they are set as 3, 10, and 20, respectively.

In this way, a single scale of the high-level features is obtained in each approach. After data processing with multiple residual connected blocks, the learned features are concatenated, and further connected with a fully connected layer for information aggregation. Therefore, the final features are in multiple scales and hold richer information from the raw data.

3.3. Deep Neural Network Structure. In this study, a deep convolutional neural network structure is used, with the residual modules and the multi-scale feature extraction method. In the proposed framework, the raw measured data are directly used as the input of the deep neural network, which means no prior expertise on the signal processing is required, which largely facilitates the practical utilization of the proposed method in the real scenarios.

Specifically, the neural network architecture is shown in Figure 2. The proposed model consists of multiple residual connected blocks. Each residual connected block typically has two convolutional layers with multiple filters of different sizes. The feature extraction scheme in three scales is generally considered. Correspondingly, three sizes of the convolutional operation are adopted in different feature extraction modules.

After feature extraction of two residual blocks in each module, the learned high-level features of different modules are concatenated for information fusion. Afterward, one fully-connected layer with 128 neurons are used, as well as the final fully-connected layer. Each neuron in the last fully-connected layer represents the predicted confidence value for each class. The softmax function at the end of the structure interprets the confidence values into the probabilities.

In the practical implementations, zero-padding operation in the convolutional layers is adopted to keep the feature map dimension unchanged. The max-pooling is also utilized in the deep model for accelerating the training process and

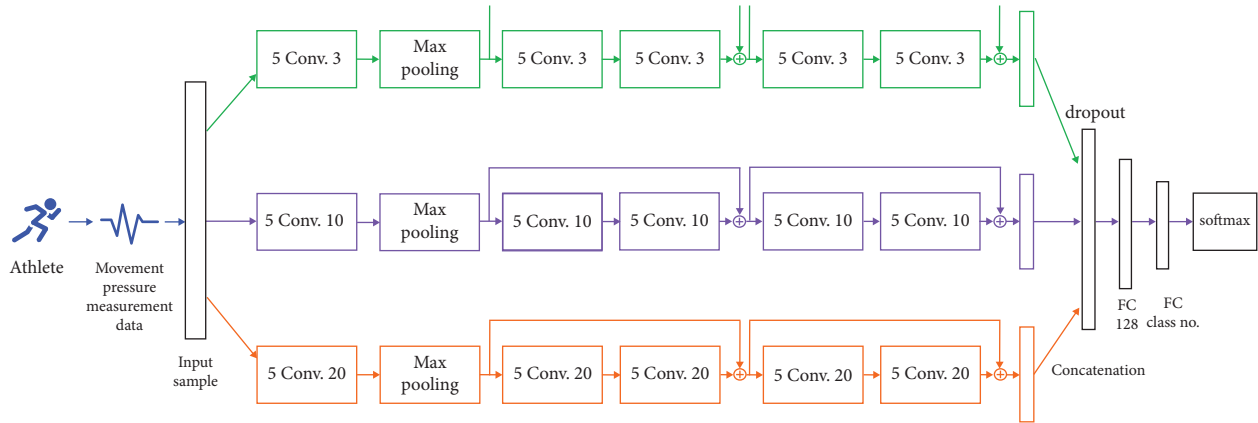


FIGURE 2: Architecture of the proposed deep multi-scale residual connected neural network model.

obtaining the significant features. Throughout the deep neural network, the leaky rectified linear unit (Leaky ReLU) activation function is adopted after the layers, which are generally stable with respect to the gradient vanishing or gradient diffusion problems and can lead to better performance. The popular cross-entropy loss function is utilized for optimization of the neural network model [48]. The back-propagation algorithm is applied for the specific changes of the model coefficients in each optimization iteration. The widely used Adam optimization method is employed for model training.

3.4. General Implementation. Figure 3 shows the flowchart of the proposed deep multi-scale residual connected model. First, the measured time series raw data are prepared into multiple samples. Specifically, in this study, the movement pressure data in two directions are used, i.e., x and y directions. Therefore, the raw data have two dimensions. The sample dimension in one direction can be defined as N_{in} , and we can prepare the samples accordingly with dimension $[2, N_{in}]$. The raw data can be directly used as the model inputs, and no prior knowledge on signal processing is needed, which shows that the applicability of the proposed method in the real scenarios is strong.

Next, with respect to the specific dataset information, the proposed deep multi-scale residual connected neural network architecture is established, and the detailed configurations are determined, including the number of neurons in the hidden layers, number of convolutional filters, and so on. In order to start the model training process, the data samples are fed into the network. Through multiple layers of feature extraction, high-level representations are obtained, which are used for the final classification. Back-propagation algorithm is used for the updates of the model parameters.

Afterward, when the model training process is finished, the testing samples are fed into the deep neural network to test the model performance with respect to the unseen data.

4. Experimental Study

4.1. Dataset and Task Description. In this study, a real athlete balance control ability evaluation dataset is used for the

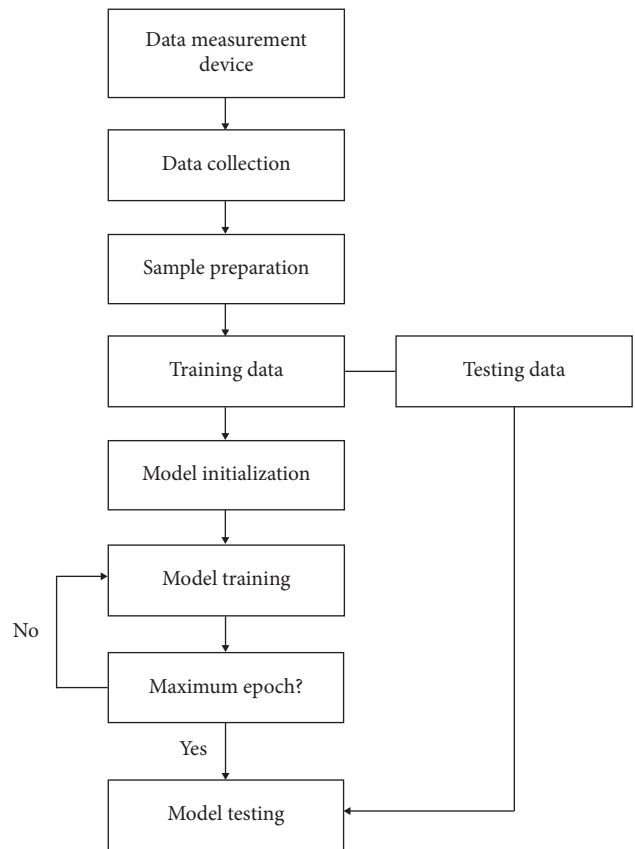


FIGURE 3: The flowchart of the proposed method in athlete balance control ability evaluation.

validation of the proposed method. Specifically, multiple freestyle skiing aerials athletes of different balance control levels are asked to stand still on a balance meter under feet. The area of the balance meter is $65\text{ cm} \times 40\text{ cm}$, and the balance meter can collect the movement pressure data in the anteroposterior and mediolateral directions.

Three levels in balance control of freestyle skiing aerials athletes are considered, which are denoted as high-level (H), medium-level (M), and normal people (N), respectively. Each level includes two athletes, who are represented by numbers of #1 and #2, respectively. The athletes are required

TABLE 1: Information of the athlete movement pressure measurement dataset used in this study.

Athlete level	No. of athletes	Code names	Sampling frequency
H (High-level athlete)	2	H#1, H#2	100 Hz
M (Medium-level athlete)	2	M#1, M#2	100 Hz
N (Normal people)	2	N#1, N#2	100 Hz

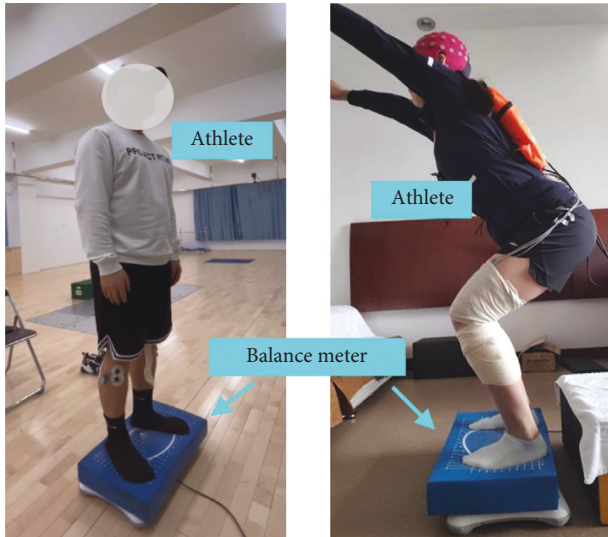


FIGURE 4: The scenarios of the freestyle skiing aerials athlete movement pressure data collection experiments.

to keep balance at their best when they are standing on the balance meter. Their upper bodies are supposed to be stationary, and the noise of the environment is kept at the minimum level. The athletes use two of their feet for standing with their eyes closed to focus on the data measurement. The movement pressure data sampling frequency is 100 Hz. Table 1 presents the information of the dataset used in this study, and Figure 4 shows the scenario of the experiment.

In this study, different athlete balance control ability prediction tasks are considered in order to fully examine the performance of the proposed method. Specifically, with respect to the dataset, four tasks are implemented, where different training and testing data are used. The tasks are demonstrated in Table 2. Different athletes in each levels are used for validation, which cover a wide range of the experimental settings and provide fair evaluations of the performance of the proposed method.

4.2. Model Establishment. In this study, mini-batch data samples are used to implement the stochastic gradient descent (SGD) optimization method for updating the deep neural network parameters. In each epoch of the training process, the training data samples are divided into different mini-batches in a random manner. Eight samples are included in each mini-batch with the corresponding label information.

Afterward, the deep neural network parameters are updated with the popular cross-entropy loss function with respect to each mini-batch. It is worth noting that the dimension of the data samples plays an important role in the model performance. Larger dimension indicates more

TABLE 2: Information of different athlete balance control ability evaluation tasks used in this study.

Task name	Concerned Athletes	No. of training Samples	No. of testing Samples
T1	H#1, M#1, N#1	1200	600
T2	H#2, M#2, N#2	1200	600
T3	H#1, M#2, N#1	1200	600
T4	H#2, M#1, N#2	1200	600

TABLE 3: Parameters of the proposed method used in this study.

Parameter	Value	Parameter	Value
Batch size	8	Sample dimension	200 * 2
Epoch number	100	Convolutional filter size	3, 10, 20
Learning rate	$1 * 10^{-4}$		

information is included in each sample. However, higher computational burden usually exists. Therefore, this is generally a trade-off in the practical applications.

The deep neural network model architecture is shown in Figure 2. The model performance can be affected by some key factors, such as the convolutional filter size and number. Those will be further investigated in the following sections in this study. Specifically, for the experiments, the parameters used in the proposed method are listed in Table 3, which are selected based on the performances on the validation data in this case.

4.3. Compared Approaches. The proposed deep multi-scale residual connected neural network model offers a new perspective for big data-driven intelligent athlete balance control performance evaluation. In this study, similar methods in the existing literature are also implemented for comparisons, in order to examine the effectiveness and superiority of the proposed methodology. Specifically, the following approaches are considered, which cover a wide range of popular techniques for data-driven studies.

4.3.1. NN. The basic neural network (NN) model is firstly considered, which follows a typical pattern for neuron connections [23]. Specifically, a multi-layer perceptron structure is used, which includes one hidden layer with 1000 neurons. Similar configurations are used as the proposed method, such as the Leaky ReLU activation function and dropout operation.

4.3.2. DNN. The deep neural network (DNN) is an extension of the basic neural network structure [49]. Three hidden layers are considered in the DNN method in this study, which consists of 1000, 1000, and 500 neurons, respectively. Similarly, the Leaky ReLU activation function is also employed, as well as the dropout technique.

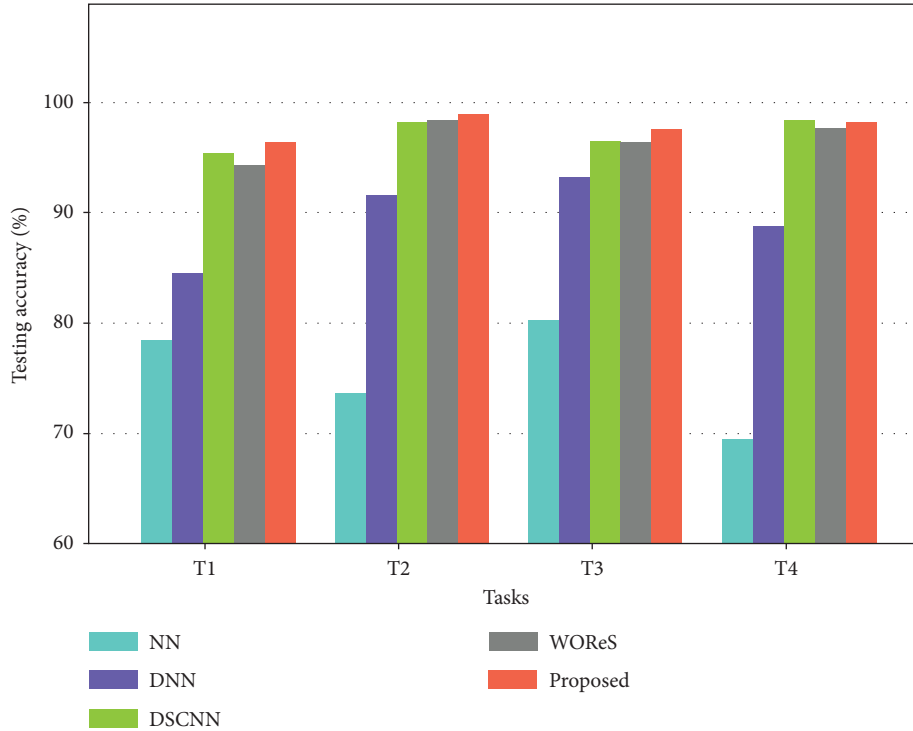


FIGURE 5: The experimental results of different compared methods in different athlete balance control ability evaluation tasks.

4.3.3. DSCNN. The deep single-scale convolutional neural network (DSCNN) method is implemented [50], which share the similar architecture with the proposed method, except for the multi-scale feature extraction scheme. Specifically, only one data processing approach is considered in the network. Correspondingly, one convolutional filter size is employed for the feature extraction. No feature concatenation is used at the fully-connected layers. The other settings are similar with the proposed method.

4.3.4. WORes. The WORes method represents the deep multi-scale convolutional neural network architecture, which does not have the residual connected schemes [40]. Specifically, the short cut connections between the convolutional layers are removed from the proposed method. This approach is a comparison to show the benefits of the proposed residual connected scheme.

With respect to all the compared methods in this study, the cross-entropy loss function is used for classification of the athlete balance control performance. The Adam optimization method is adopted for the model updates with the mini-batch data sample selections. The same learning rate is used as the proposed method.

5. Experimental Results and Performance Analysis

In this section, the experimental results of the proposed method on different athlete balance control ability evaluation tasks are presented, as well as the results of different compared methods. Ablation studies are also extensively

carried out to evaluate the influence of different key parameters of the proposed method on the model performance. In order to provide fair results and comparisons, each experiment is implemented for three times, and average results are presented.

Figure 5 shows the general experimental results using different methods in different tasks. It can be observed that in general, the neural network-based methods are able to achieve good evaluation results, and the testing accuracies are high. The testing accuracies of the basic NN method are not competitive in different tasks, and less than 80% accuracies are obtained. This indicates that the shallow network structure cannot well capture the underlying pattern of the massive data. The DNN method achieves significantly higher testing accuracies in different tasks, and the accuracies are basically higher than 90%. The results show that the deep architecture can well learn the highly nonlinear relationship between the movement pressure measurement data and the athlete balance control ability. The DSCNN and WORes methods are quite competitive in this problem, and the testing accuracies in different cases are mostly higher than 95%. However, the optimal performance is generally achieved by the proposed deep multi-scale residual connected model. Close to 100% testing accuracies in different tasks can be obtained. Noticeable improvements can be observed compared with the DSCNN and WORes methods. This implies that the proposed multi-scale feature learning scheme and residual connected structure can well enhance the learning performance of the deep neural network architecture, and they are well suited for the athlete balance control ability evaluation problem by processing the time-series pressure data.

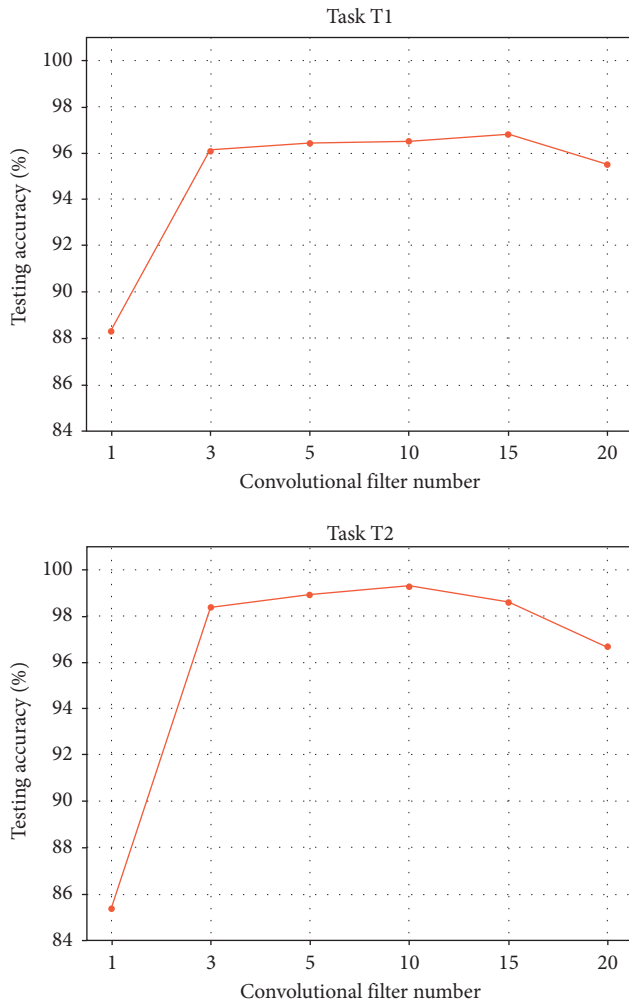


FIGURE 6: The influence of the number of convolutional filters on the model performance in the tasks T1 and T2.

5.1. Effect of Convolutional Filter Number. The number of convolutional filters in the layers throughout the deep neural network plays an important role in affecting the model performance. Fewer convolutional filters are generally less effective in learning the complicated patterns from the data, and more convolutional filters will basically lead to better performance with enhanced learning capacity. However, the overfitting issue may occur since larger model architecture and more parameters are included. In this section, this issue is investigated, and the effects of the convolutional filter number on the model performance are presented in Figure 6. The tasks T1 and T2 are used for investigation.

It can be observed that in general, the influence of the convolutional filter number on the testing accuracies is not quite significant when the number is not very small. When only one convolutional filter is used, remarkably low testing accuracies are obtained, which are lower than 90%. However, when more convolutional filters are applied, the testing accuracies are generally stable and higher than 95%. When 20 convolutional filters are employed, slight performance drops are observed. Nonetheless, this does not have noticeable influence on the general model performance.

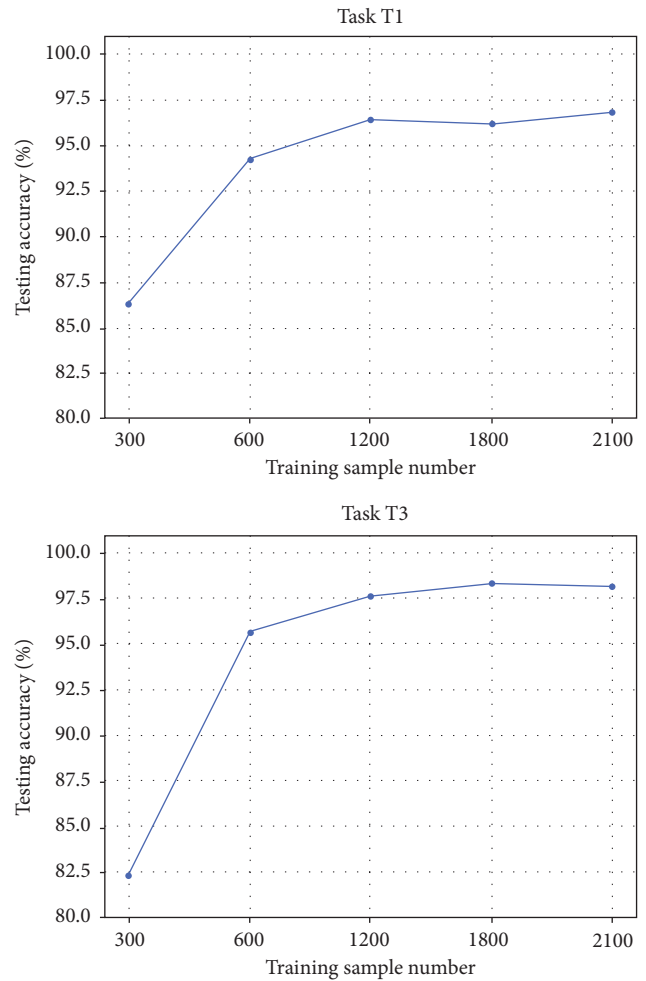


FIGURE 7: The influence of the sample number on the model performance in the tasks T1 and T3.

Therefore, when the number is not too small with a reasonable value, promising results can be basically achieved.

5.2. Effect of Sample Number. In this section, the effects of the sample number on the model performance are investigated. The number of training samples is also an important parameter in the data-driven methods. Generally, more training samples lead to better performances. However, since the data are usually expensive in different areas, it is always preferred to achieve good performance with minimum data. The experimental results are presented in Figure 7. The tasks T1 and T3 are focused on in this section.

It is noted that the experimental results are basically in accordance with our understanding in the literature. When 300 training samples are used, lower testing accuracies are obtained in different tasks, which are lower than 87%. When more training samples are employed, the results significantly become better and higher than 92% testing accuracies are basically obtained. When the sample number is larger than 600, small fluctuations of the testing performance are observed. However, the performances are generally stable with respect to different sample numbers. It is also noted that 600

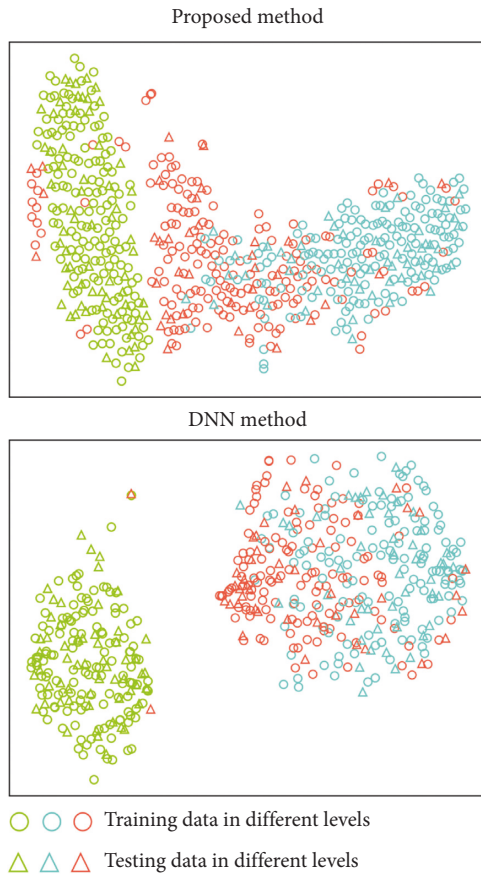


FIGURE 8: The visualization results of the learned features by different methods in task T1.

training samples are mostly sufficient for building the deep neural network model for this problem, which can be considered as the minimum number that the model requires.

5.3. Visualization of Learned Representation. In this section, the learned features by the deep neural network models are visualized to show the effectiveness of the methods. Specifically, the high-level representations of the samples at the last fully-connected layer are considered. The t-SNE method is adopted for dimension reduction of the learned high-dimensional features. Two new dimension can be obtained and plotted for visualizations. The results in the tasks T1 and T4 are shown in Figures 8 and 9 respectively.

It can be observed that using the proposed deep multi-scale residual connected neural network method, different classes are more separated with respect to the learned features. Limited overlappings between different classes are found, which validates that the proposed method can achieve high testing accuracies for the classification tasks. The DNN method is less competitive in the cases. Noticeable overlappings between different classes are observed in the learned feature sub-space, and some data samples are also located outside the clusters of their own classes. This shows that the DNN method is less effective than the proposed method in the tasks. It should be pointed out that the NN method is far less effective in the case studies, and

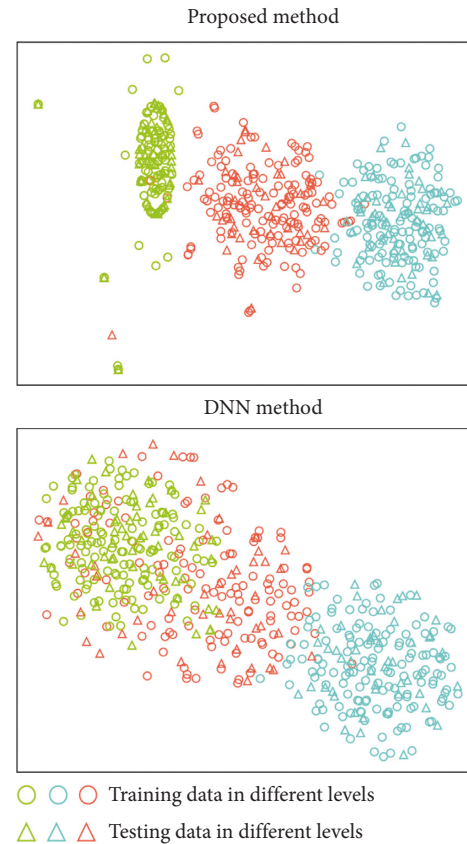


FIGURE 9: The visualization results of the learned features by different methods in task T4.

the visualization results do not carry sufficient information for demonstrating the effects. The results in this section validate the effectiveness of the proposed method in an intuitive way, which shows that the proposed method is quite promising for automatic athlete balance control ability evaluation.

6. Conclusion

In this study, a deep multi-scale residual connected neural network model for intelligent athlete balance control ability evaluation. The time-series pressure measurement data under-feet are processed and analyzed. The raw data are directly used as the model input for automatic evaluations. No prior knowledge on signal processing is needed, which makes it easy for real applications. A multi-scale feature extraction scheme is proposed, which utilize the learned features from different types of convolutional filters. The information fusion of the learned features further enhances the model training ability. The proposed residual connected blocks can effectively increase the model training efficiency while keeping the training quality. This is well suited for the deep neural network architecture and can be readily applied in different network structures. Experiments on the real athlete under-feet pressure measurement data are carried out for validations. The results show that the proposed method is promising for intelligent evaluations of the athlete

balance control abilities, and offers a new perspective in mining athlete measurement data.

The advantage of the proposed method lies in the end-to-end modeling structure, which makes the balance control ability evaluation task more straight-forward to implement. On the other hand, despite the promising results, it should be pointed out that main drawback of the proposed method lies in the structure of the neural network model, since three network approaches are considered in the model, which is a little complex for the data-driven model. Further research works will be carried out on the optimization of the deep neural network architecture while retaining the model performance.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors stated that there are no conflicts of interest.

Acknowledgments

This study was funded by the key R&D plan of China for Winter Olympics (No. 2021YFF0306401), the Key Special Project of the National Key Research and Development Program "Technical Winter Olympics" (2018YFF0300502 and 2021YFF0306400), and the Key Research Program of Liaoning Province (2020JH2/10300112).

References

- [1] N. Snyder and M. Cinelli, "Comparing balance control between soccer players and non-athletes during a dynamic lower limb reaching task," *Research Quarterly for Exercise & Sport*, vol. 91, no. 1, pp. 166–171, 2020.
- [2] A. Andreeva, A. Melnikov, D. Skvortsov et al., "Postural stability in athletes: the role of age, sex, performance level, and athlete shoe features," *Sports*, vol. 8, no. 6, p. 89, 2020.
- [3] J. A. Diekfuss, C. K. Rhea, R. J. Schmitz et al., "The influence of attentional focus on balance control over seven days of training," *Journal of Motor Behavior*, vol. 51, no. 3, pp. 281–292, 2019.
- [4] S. M. Bruijn and J. H. van Dieën, "Control of human gait stability through foot placement," *Journal of The Royal Society Interface*, vol. 15, no. 143, Article ID 20170816, 2018.
- [5] C. Yang, K. Zhou, and J. Liu, "Supergraph: spatial-temporal graph-based feature extraction for rotating machinery diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 4, pp. 4167–4176, 2022.
- [6] L. Li, Z. Luo, F. He, K. Sun, and X. Yan, "An improved partial similitude method for dynamic characteristic of rotor systems based on levenberg-marquardt method," *Mechanical Systems and Signal Processing*, vol. 165, Article ID 108405, 2022.
- [7] J. Liu, K. Zhou, C. Yang, and G. Lu, "Imbalanced fault diagnosis of rotating machinery using autoencoder-based supergraph feature learning," *Frontiers of Mechanical Engineering*, vol. 16, no. 4, pp. 829–839, 2021.
- [8] M. Hilbert, W. Smith, and R. Randall, "The effect of signal propagation delay on the measured vibration in planetary gearboxes," *Journal of Dynamics, Monitoring and Diagnostics*, vol. 1, no. 1, pp. 9–18, 2022.
- [9] K. Zhou, C. Yang, J. Liu, and Q. Xu, "Dynamic graph-based feature learning with few edges considering noisy samples for rotating machinery fault diagnosis," *IEEE Transactions on Industrial Electronics*, p. 1, 2021.
- [10] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Federated learning for machinery fault diagnosis with dynamic validation and self-supervision," *Knowledge-Based Systems*, vol. 213, Article ID 106679, 2021.
- [11] V. T. Tran, B.-S. Yang, F. Gu, and A. Ball, "Thermal image enhancement using bi-dimensional empirical mode decomposition in combination with relevance vector machine for rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 38, no. 2, pp. 601–614, 2013.
- [12] Y. Xu, X. Tang, X. Tang et al., "Orthogonal on-rotor sensing vibrations for condition monitoring of rotating machines," *Journal of Dynamics, Monitoring and Diagnostics*, vol. 1, no. 1, pp. 29–36, 2021.
- [13] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Open set domain adaptation in machinery fault diagnostics using instance-level weighted adversarial learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 11, pp. 7445–7455, 2021.
- [14] I. Yesilyurt, A. Dalkiran, O. Yesil, and O. Mustak, "Scalogram-based instantaneous features of acoustic emission in grinding burn detection," *Journal of Dynamics, Monitoring and Diagnostics*, vol. 1, no. 1, pp. 19–28, 2021.
- [15] W. Wang, Y. Lei, T. Yan, N. Li, and A. Nandi, "Residual convolution long short-term memory network for machines remaining useful life prediction and uncertainty quantification," *Journal of Dynamics, Monitoring and Diagnostics*, vol. 1, no. 1, pp. 2–8, 2021.
- [16] S. Siahpour, X. Li, and J. Lee, "A novel transfer learning approach in remaining useful life prediction for incomplete dataset," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.
- [17] B. Sun and K. Saenko, "Deep CORAL: correlation alignment for deep domain adaptation," in *Proceedings of ECCV*, pp. 443–450, Springer, Cham, Manhattan, NY, USA, 2016.
- [18] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Transfer learning using deep representation regularization in remaining useful life prediction across operating conditions," *Reliability Engineering & System Safety*, vol. 211, Article ID 107556, 2021.
- [19] J. Weston, F. Ratle, H. Mobahi, and R. Collobert, "Deep learning via semi-supervised embedding," in *Neural Networks: Tricks of the Trade*, G. Montavon, G. B. Orr, and K.-R. Müller, Eds., pp. 639–655, Springer Berlin Heidelberg, Berlin, Heidelberg, Second Edition, 2012.
- [20] C. Szegedy, W. Liu, J. Yangqing et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, IEEE, Boston, MA, 7 June 2015.
- [21] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, "Degradation alignment in remaining useful life prediction using deep cycle-consistent learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2021.
- [22] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Universal domain adaptation in fault diagnostics with hybrid weighted deep adversarial learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 7957–7967, 2021.
- [23] A. Ainapure, S. Siahpour, X. Li, F. Majid, and J. Lee, "Intelligent robust cross-domain fault diagnostic method for rotating machines using noisy condition labels," *Mathematics*, vol. 10, no. 3, p. 455, 2022.

- [24] A. A. Adegun, S. Viriri, and R. O. Ogundokun, "Deep learning approach for medical image analysis," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 6215281, 9 pages, 2021.
- [25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Proceedings of 26th Annual Conference on Neural Information Processing Systems*, vol. 2, pp. 1097–1105, 2012.
- [26] J. T. Zhou, S. J. Pan, I. W. Tsang, and Y. Yan, "Hybrid heterogeneous transfer learning through deep learning," in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp. 2213–2219, AAAI Press, Vancouver, Canada, 27 July 2014.
- [27] G. Hinton, L. Deng, D. Yu et al., "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [28] J. Liu, C. Zhang, and X. Jiang, "Imbalanced fault diagnosis of rolling bearing using improved MsR-GAN and feature enhancement-driven CapsNet," *Mechanical Systems and Signal Processing*, vol. 168, Article ID 108664, 2022.
- [29] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Journal of Machine Learning Research*, vol. 9, pp. 249–256, 2010.
- [30] Z. He, H. Shao, X. Zhong, and X. Zhao, "Ensemble transfer cnns driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions," *Knowledge-Based Systems*, vol. 207, Article ID 106396, 2020.
- [31] Y. Feng, H. Zhang, W. Hao, and G. Chen, "Joint extraction of entities and relations using reinforcement learning and deep learning," *Computational Intelligence and Neuroscience*, vol. 2017, Article ID 7643065, 11 pages, 2017.
- [32] W. Zhang, X. Li, and X. Li, "Deep learning-based prognostic approach for lithium-ion batteries with adaptive time-series prediction and on-line validation," *Measurement*, vol. 164, Article ID 108052, 2020.
- [33] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: a systematic literature review: 2005–2019," *Applied Soft Computing*, vol. 90, Article ID 106181, 2020.
- [34] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 917–963, 2019.
- [35] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, "Partial transfer learning in machinery cross-domain fault diagnostics using class-weighted adversarial networks," *Neural Networks*, vol. 129, pp. 313–322, 2020.
- [36] H. Yang, X. Li, and W. Zhang, "Interpretability of deep convolutional neural networks on rolling bearing fault diagnosis," *Measurement Science and Technology*, vol. 33, no. 5, Article ID 055005, 2022.
- [37] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," *Proceedings of 32nd International Conference on Machine Learning*, vol. 37, pp. 97–105, 2015.
- [38] F. Miao, B. Wen, Z. Hu et al., "Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques," *Artificial Intelligence in Medicine*, vol. 108, Article ID 101919, 2020.
- [39] C. Mataczyński, A. Kazimierska, A. Uryga, M. Burzyńska, A. Rusiecki, and M. Kaspruwicz, "End-to-end automatic morphological classification of intracranial pressure pulse waveforms using deep learning," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 2, pp. 494–504, 2022.
- [40] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015, <https://arxiv.org/abs/1512.03385>.
- [41] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training Very Deep Networks," in *Proceedings of the 28th International Conference On Neural Information Processing Systems*, pp. 2377–2385, Mit Press, Montreal, Canada, 7 December 2015.
- [42] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, vol. 1, pp. 448–456, jmlr.org, Lille, France, 6 July 2015.
- [43] W. Zhang and X. Li, "Federated transfer learning for intelligent fault diagnostics using deep adversarial networks with data privacy," *IEEE*, vol. 27, no. 1, 2021.
- [44] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proceedings of the European Conference on Computer Vision*, pp. 630–645, Springer, Amsterdam, The Netherlands, 8 October 2016.
- [45] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [46] K. B. Lee, S. Cheon, and C. O. Kim, "A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 30, no. 2, pp. 135–142, 2017.
- [47] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017.
- [48] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," 2015, <https://arxiv.org/abs/1503.02531>.
- [49] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: a review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, Article ID 106587, 2020.
- [50] W. Ren, J. Pan, H. Zhang, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks with holistic edges," *International Journal of Computer Vision*, vol. 128, no. 1, pp. 240–259, 2019.

Review Article

Potential Future Directions in Optimization of Students' Performance Prediction System

Sadique Ahmad ¹, Mohammed A. El-Affendi ¹, M. Shahid Anwar ²
and Rizwan Iqbal ³

¹EIAS: Data Science and Blockchain Laboratory, College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia

²Department of Artificial Intelligence and Software, Gachon University, Seongnam, Republic of Korea

³Department of Computer Engineering, Bahria University, Karachi Campus, Karachi, Pakistan

Correspondence should be addressed to Sadique Ahmad; sadiqueahmad.bukc@bahria.edu.pk

Received 19 February 2022; Accepted 26 March 2022; Published 17 May 2022

Academic Editor: Zhongxu Hu

Copyright © 2022 Sadique Ahmad et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Previous studies widely report the optimization of performance predictions to highlight at-risk students and advance the achievement of excellent students. They also have contributions that overlap different fields of research. On the one hand, they have insightful psychological studies, data mining discoveries, and data analysis findings. On the other hand, they produce a variety of performance prediction approaches to assess students' performance during cognitive tasks. However, the synchronization between these studies is still a black box that increases prediction systems' dependency on real-world datasets. It also delays the mathematical modeling of students' emotional attributes. This review paper performs an insightful analysis and thorough literature-based survey to draw a comprehensive picture of potential challenges and prior contributions. The review consists of 1497 publications from 1990 to 2022 (32 years), which reported various opportunities for future performance prediction researchers. First, it evaluates psychological studies, data analysis results, and data mining findings to provide a general picture of the statistical association among students' performance and various influential factors. Second, it critically evaluates new students' performance prediction techniques, modifications in existing techniques, and comprehensive studies based on the comparative analysis. Lastly, future directions and potential pilot projects based on the assumption-based dataset are highlighted to optimize the existing performance prediction systems.

1. Introduction

Over the past few decades, students' performance has been predicted while evaluating the influence of different factors, such as emotional attributes, family attributes, study schedule, institutional attributes, and students' scores in assignments, quizzes, and final examinations [1–5]. Such systems provide useful applications to a wide area in academia, i.e., students' success and failure estimation due to influential factors [6–10]. This study splits the earlier contributions into two groups. The first group consists of insightful psychological studies, data mining discoveries, and data analysis findings that indirectly contribute to the optimization of students' performance prediction systems. The

second group reports the optimization of existing prediction systems based on the findings of the first group. However, the extensive synchronization between the two groups is still a black box that ultimately increases students' performance prediction systems' dependency on a real-world dataset. Such synchronization can provide useful ideas during the optimization and data collection process. It also paves the way for an assumption-based dataset to prove the viability of pilot project implementations that will speed up modeling students' emotional attributes.

This review paper conducts an insightful study and literature-based survey to draw a comprehensive picture of the prior studies on student performance analysis and prediction. The review consists of 1497 articles from 1990 to

2022 (32 years), which reported various information for future researchers:

1. It explores and lists the research fields' contributions focusing on students' performance optimization. Psychology and data analysis fields pave the way for effective solutions to the problems of data deficiency. They provide qualitative findings that can be used for creating an assumption-based dataset for pilot project implementation.
2. It thoroughly considered new and modified algorithms that predict students' performance. Also, a comparative analysis was performed between the existing students' performance prediction approaches to provide better recommendations for optimization.
3. The study delivers a comprehensive picture of potential challenges and research direction for future researchers. The review also shows that very few contributions have mathematically modeled emotional attributes.

The remaining sections of this review are as follows: Section 2 gives a detailed literature review. Section 3 elaborates the review methodology, and Section 4 produces data evaluation. Section 5 presents future challenges, and Section 6 concludes the study.

2. Literature Review

Students' performance prediction systems have enormous applications in academia, such as predicting at-risk students, course recommendation, and basic counseling against negative emotions, highlighting the influence of institutional attributes, family factors, etc. [11–14]. It is also needed to advance the academic achievement of excellent students [15–19]. Prior studies deliver qualitative and quantitative results in extending students' performance evaluation and calculation, and highlighting the factors that influence the performance [20–25]. For a few decades, psychology, data mining, cognitive computing, and data analysis fields directly or indirectly contributed to the optimization of students' performance prediction systems [26–34]. Therefore, related work is split into the following subsections.

2.1. Contributions of Psychology. Psychological studies results manifest that students' performance is easily influenced via emotional attributes, such as frustration, anxiety, stress, over expectation of parents, and parents' relationship [35–37]. The results provide correlations statistics among emotional factors and the expected performance of students in cognitive activity, such as attempting the examination, quizzes, assignments, class activities, and extracurricular activities. These particular emotional factors can negatively and positively impact the students' performance. In such a situation, emotional severity, family attributes, and institutional factors play a crucial role in influencing performance [38–40]. It shows that performance is always very sensitive and affected by the individuals' surroundings.

2.2. Contributions of Data Mining. The data mining evaluates the relationship among various students' factors, such as the role of emotional factors, family attributes, institutional factors, and class performance. Such studies provide good opportunities for accurate estimations of expected students' performance [41–47]. The meaningful patterns always produce good directions for further exploration. The previous articles lack coordination between the students' influential attributes and academic performance. The literature lacks accurate techniques to simulate students' performance due to the insufficient synchronization and coordination among earlier studies on students' factors. The mathematical modulation of students' performance needs to formulate the function of student factors. However, it is inspiring to closely examine the quantitative influence of several student factors on academic achievements. The earlier studies show that emotional, family, study schedule, and institutional attributes are the significant factors that can easily influence students' academic performance in any critical cognitive activity. Prior studies illustrate that educational data mining practices contribute to students' factor evaluation process and performance prediction. Institutional factors involve teaching methodology, engaging students in the classroom, and the vision of instruction. According to literature studies, teachers play an active role in institutional attributes influencing students' performance. They provide administrative assistance and assistance in ensuring discipline [48–57].

2.3. Synchronization among Existing Studies. Accurate performance prediction needs to examine students' factors beyond the computer science framework. The literature studies are still limited in finding an authentic and extendable approach that overlaps psychology, data mining, data analysis, and cognitive research. Articles have various solutions to predict students' performance using different techniques that could have the potential to be escalated to more general problems of predicting student performance [58–62]. The primary objective of the current review attempt is to efficiently explore the relationship between students' factors (as mentioned earlier) and their performance. Therefore, the literature is studied with the selected students' attributes (emotional, family, study schedule, and institutional) and effects. Articles show that most students do not participate in extracurricular activities, believing extra activities would negatively affect their academic achievements. Earlier studies also focus on predicting college students' performance by considering all the important aspects. They delivered a prediction system to estimate performance by assisting the university in selecting each candidate using past academic records of students granted admissions [29, 63–69]. Such efforts show that earlier studies contribute to decreasing the number of at-risk students and advancing the performance of excellent students.

Literature also attempted to perform a survey on classroom learning in different environments. It analyzes various aspects and factors influencing (positively or negatively) performance in a classroom that interfere with

learning. This paper presents a systematic review of numerous studies on students' performance in classroom learning. For a few decades, the research has produced numerous results in students' performance evaluation; however, the education system needs a complete and detailed performance prediction system that can ensure interaction and coordination between the aforementioned students' factors. Literature studies delivered various contributions, such as the proposal of an innovative model that targets modifying learning sustainability through smart education applications and regression and correlation among students' factors, and logistic regression analyses generated that being female, first-semester GPA, number of courses per regular semester, and number of courses per summer semester were imperative predictors of baccalaureate degree achievement [70–77].

2.4. Existing Models and Performance Prediction System Optimization. Studies have focused on applying artificial neural networks to predict performance in different environments. Articles are also saturated with deep learning techniques that deliver prediction and highlight at-risk students. Few other technologies provide opportunities to accurately evaluate the performance and reduce the failure rates [78–82]. It also helps in counseling students in alarming situations that can positively impact their academic achievements, i.e., COVID-19. Thus, during the literature survey, we have found many students' prediction systems which are interesting; nevertheless, they are failed to mathematical model emotional attributes and synchronized them with institutional attributes, study schedules, and family attributes [31, 37, 83–98]. The objective of this study is to identify the relationship between extracurricular activities and students' performances.

The articles deliver many results on the effects of influential students' factors. This study explores performance prediction beyond the scope of computer science and machine learning.

2.5. Related Performance Prediction Methods. As discussed earlier, many studies solved meaningful challenges in students' performance prediction area of research for a few decades. The earlier studies have many contributions in the form of neural works, recommendation systems, course recommendations, and students' performance evaluation systems [87, 99–109]. The prior studies demonstrate comprehensive work on students' performance prediction systems that use information obtained during the interaction of students with the institutional attributes. To mathematically consider the expected actions of a students' factors, such information provides proper guidelines. The significant characteristic is the identical structure of the information processing system of students, which can be replicated to construct a learning algorithm (cognitive architecture). Literature studies are flooded with many findings that primarily contribute to prediction algorithms and mathematical models; nevertheless, modeling the relationship

between students' emotions (frustration, stress, etc.) and students' performance is very little focused [110–115].

Also, the published studies on modeling emotion are not extendable toward a matured prediction system. So, the dire need is to assess the main framework of existing prediction algorithms. Exploring the qualitative results of psychological studies and data analysis discoveries is needed to estimate students' performance. It will also help in the iterative calculation of emotional influence on performance during critical cognitive activities.

Extraordinary academic performance is only possible with excellent cognitive skills. Such skills are needed to accomplish any task requiring problem-solving approaches, reasoning, and memory management. However, with inadequate cognitive abilities, an individual cannot achieve an excellent score in various cognitive tasks, i.e., assignments, quizzes, and written examinations. They require students to process new information, organize learning, and retrieve that data (from memory) for later use. So, predicting performance while calculating the intense impact of various groups of factors is crucial not only for tutors to ensure effective teaching methodology but also for students' achievements and effective academic policies. Earlier studies have delivered many approaches that predict students' performance; nevertheless, they have paved the way for new challenges for effective educational systems. The skills levels of students are changing as they learn and forget. The educational system needs such a system that can manage the students' dynamic behavior during cognitive activities.

Other studies have described the students' personality traits and the essential characteristics of personality. Results reveal that performance prediction design can be broken down into subsections that are more realistic in comparison to other techniques. It also paves the way for the development of performance prediction architectures which were easy to understand. The current review illustrates that the performance prediction system needs to coordinate among prediction architecture, psychological experiments, semantical investigations, statistical analysis, and mathematical formulation. This work provided an outstanding opportunity for researchers belonging to performance prediction, bioinformatics, data mining, and data integration.

3. Review Methodology

The review process is started from the initial screening within the scope of the current attempt. As elaborated earlier, this review focused on recent and state-of-the-art contributions to student performance prediction. Figure 1 illustrates the methodology of the review. We have divided the complete review process into the following sections.

3.1. Review Process. This study reviews the earlier studies thoroughly based on the procedures prescribed by Petersen et al. [116] and Keele [117]. The methodology is adopted from Keele while the study mapping method is copied from Petersen et al. The review process is initiated with the

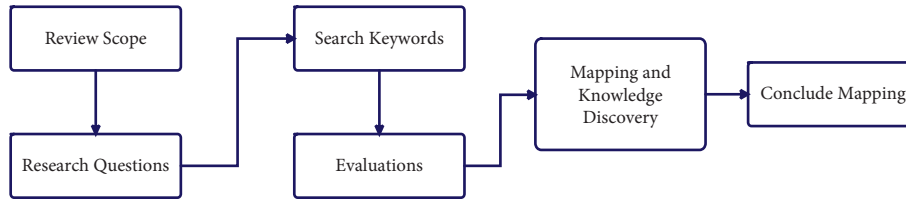


FIGURE 1: Framework illustrated the main modules of the current review process.

TABLE 1: Result obtained from the Google Scholar during keyword searching.

No	Keywords	IEEE	ACM	Springer	MDPI	Hindawi	Elsevier	Wiley	Others	Total
1	Student performance prediction	14	13	8	5	12	7	7	7	73
2	Student performance and negative emotions	9	8	4	3	12	9	7	6	58
3	Student emotional factors	9	7	1	2	2	4	6	2	33
4	Work experience and student performance	20	10	4	4	2	5	3	4	52
5	Student biological factor and academics	11	5	6	2	3	2	5	10	44
6	Student academic achievements prediction	11	7	8	6	3	5	1	6	47
7	Student frustration	4	2	5	2	2	2	3	7	27
8	Student performance and frustration	3	7	7	6	4	3	4	7	41
9	Student frustration severity	5	7	9	5	0	8	13	13	60
10	At-risk student prediction	6	9	6	11	11	7	7	11	68
11	At-risk student cognitive skills	3	5	3	8	4	3	8	2	36
12	Cognitive skills prediction	8	6	4	7	4	5	10	10	54
13	Emotional impact on student performance	6	9	7	5	2	9	7	5	50
14	Family impact on student achievements	7	2	5	9	8	4	11	5	51
15	Student anxiety	9	7	2	11	6	7	2	5	49
16	Student stress	2	9	8	11	9	7	2	7	55
17	Review on student performance	3	13	8	3	4	5	4	3	43
18	Student performance quantization	8	3	6	2	11	2	8	4	44
19	COVID-19, frustration, and student performance	7	5	5	3	9	8	11	4	52
20	COVID-19 and at-risk student	8	8	8	4	6	3	6	2	45
21	Impact of online classes	3	2	8	8	14	7	3	2	47
22	Online classes and student learning	7	11	2	6	9	8	10	8	61
23	Learning prediction	2	3	4	4	8	2	8	5	36
24	Student learning outcome prediction	6	5	5	4	9	6	6	9	50
25	Student performance measurement	2	12	4	4	8	4	9	7	50
26	Performance measurement algorithm	5	4	2	4	9	6	6	5	41
27	Performance prediction algorithm	11	4	5	6	6	2	9	11	54
28	Student performance evaluation algorithm	7	2	6	2	6	2	3	6	34
29	Student performance prediction algorithm	12	3	9	4	8	12	15	5	68
30	Student performance measurement algorithm	3	6	2	3	7	5	2	4	32
31	Cognitive skills prediction algorithm	3	7	5	7	4	5	6	5	42

modified procedure, which is demonstrated in Figure 1. For better understanding, the review delivers a detailed methodology of the prior work contributing to student performance prediction directly or indirectly. Moreover, the study put a list of research questions to demonstrate the main objectives. These research questions enable us to choose relevant research studies for screening and investigating the main challenges in students' performance predictions. Every research question has a list of keywords to explore the literature and learn about a particular question. These keywords are used to search publications, including peer-reviewed book chapters, conferences proceeding, and journal articles.

3.2. Research Questions

1. Q1: what are the applications of student performance prediction systems?

2. Q2: what are the factors that can optimize student performance prediction?
3. Q3: what is the intensity of research findings in the field of student performance prediction systems optimization?
4. Q4: are the findings of psychological studies, data mining, and contribution in algorithms synchronized with each other for the viability of the pilot project?
5. Q5: how synchronization and coordination of prior psychological, data mining, and algorithmic findings contribute to the effective educational system via student performance prediction algorithm.

3.3. *Searching Keywords.* The current study adopted a manual review methodology introduced by Keele [117]. The automatic review presented by Petersen et al. has a few

disadvantages [116]. (1) The automatic search is not feasible for the current review [118]. (2) The manual searching strategy gives more relevant studies. Table 1 reflects the list of keywords that have produced a variety of articles published by various publishers, i.e., IEEE, Elsevier, Springer, Hindawi, MDPI, ACM, Wiley, and others. It shows many articles, including journals, book chapters, and conference proceedings.

The keywords were searched directly on publishers' websites and Google Scholar with a default setting. We have evaluated all the articles and collected those that deliver relevant findings for further screening. Furthermore, the main factors, topics, and relevant studies, including journals, conference proceedings, and book chapters, are given below:

1. Emotional attributes
2. Family factors
3. Study Schedule
4. Institutional attributes
5. Psychology, data mining, and data analysis findings on the factors as mentioned earlier
6. Contribution of cognitive computing, deep learning, and machine learning in students' performance prediction
7. Reviews and comparison

3.4. Screening. Screening of studies is performed with the following terms and conditions:

1. The team selected the publications of the more relevant journal, conference, and book chapter.
2. Second, we have focused on the relevant title with impressive citations in Google Scholar.
3. Third, rapid reviews were performed for further evaluation and data extraction. During the rapid review, we have focused on the abstract and introduction to get some idea about the challenges, motivations, and contributions.

These three steps were performed to create a database for further information extraction and data collection.

3.5. Information Collection. Various information was extracted from the selected publications during the information collection process, which are shown in Table 1 to 4. Also, a spreadsheet was used to record the various information for further consideration of the research questions. The recorded data are shown in the tables mentioned above.

4. In-Dept Analysis

4.1. Q1: What Are the Applications of Student Performance Prediction Systems. A performance prediction system is essential to predict at-risk students to devise a solution for successful graduation and goal achievements, such as special treatment and counseling sessions. Such prediction systems are more challenging due to the significant factors affecting

students' performance. Thus, a systematic review of the literature has been performed to highlight potential issues in predicting student performance. The study also shows the contributions of previous articles beyond the scope of artificial intelligence, i.e., data mining, data analysis, and psychology techniques contributing to performance prediction. Also, this study provides an overview of prediction techniques that have been used to estimate performance. It focuses on how the predictive algorithm can be used to identify key attributes in influencing students' academic achievements. With the help of data mining and machine learning techniques in education, the study could have a more effective methodology in proposing a new prediction algorithm and modifying existing students' performance prediction systems. The primary application outcomes of students' performance prediction are given below.

4.1.1. Prediction of At-Risk Student. It is crucial to predict at-risk students and devise an effective learning environment in classrooms and laboratories. Although the literature studies are saturated with tremendous results, it is still challenging as the prediction system cannot synchronize and mathematically model emotional attributes, family issues, study schedules, and institutional attributes to develop a significant prediction system. The current review's first target is to highlight the possibilities of predicting at-risk students while coordinating between literature studies.

4.1.2. Advances the Students' Academic Achievements. The performance prediction system is essential for at-risk, average, and excellent students. The influential factors that drive academic achievement are an eternal global challenge associated with students, families, teachers, and educational policymakers. Exploring these factors benefits all those interested in developing a system for students' performance prediction worldwide. Suppose the prediction system considers a large number of influential factors. In that case, the academic achievement of excellent students can also be advanced, i.e., the prediction system could highlight problems due to various emotional, study schedules, family, and institute-related attributes.

4.1.3. Monitoring Students' Behavior. Student behavior plays a significant role in improving academic achievements, such as interaction and attitude with the teachers, seriousness, and unseriousness in the classroom. Articles of psychology and data analysis contribute to student behavior evaluation, merits, and demerits of various aspects of behavior. We need a prediction system that efficiently modulates the relationship between behavior and students' performance to highlight, monitor, and improve students' interaction and engagement in the classroom. It is also essential for the institution to devise effective controlling policies to counter and control the demerit of various behaviors. Through such a prediction system, teachers can easily guide their students in setting and achieving academic goals. A teacher can also help students understand their behavior and its impact on

others. The adverse effects of behavior can be overcome and later on monitored by supervising students. Such a system enhances the overall reputation of the institution. Other benefits include preventing early school drop-outs and building good relationships among students. According to Kennelly and Monrad [119, 120], the behavioral problem plays a key role in indicating students at risk and highlighting the individuals near to being dropped off at the institute. Therefore, employing strategies to monitor and control student behavior is extremely important for an effective educational system in a society.

4.2. Q2: What Are the Factors That Can Optimize Student Performance Prediction? The literature studies indicate that many factors influence students' performance in cognitive activities, such as quizzes, assignments, examinations, and homework. It includes family-related factors, emotional factors, gender description, and institution-related factors. A brief description of these factors is given below.

4.2.1. Family-Related Factors. The parental involvement and their particular influence are two-fold. First, the earlier studies claim that the interaction of parents positively influences performance. It enhances the academic achievements of the student in critical environments. Research results highlight that parents' friendly attitudes positively affect student performance, such as daily engagement in cognitive activities. Positive parent involvement can advance the performance, and that father or mother is the first teacher who plays the role of an enduring educator. Such research findings show that parents' positive and active role cannot be underestimated. Second, the overexpectation of parents can push children towards frustration [121, 122]. Parent mostly observes remarkable achievement on social media, so they also start demanding good grades from their children. With such pressure, students are easily frustrated, which negatively influences their academic outcome during cognitive activities, such as assignments, quizzes, and mid-and final-term examinations. So, the role of the parents should be supportive and motivational, which would help against unnecessary pressure.

To achieve a student performance prediction system, we need to consider parental involvement and the aforementioned other attributes, such as the cohabitation status of parents, the relationship among their parents, socioeconomic situation, and the number of children. Prediction systems need to quantize all these attributes to evaluate future student performance properly. If we look into literature studies, a minimal contribution can be evidenced toward mathematical modeling of student performance for a better educational system.

4.2.2. Emotional Factors. Emotional attributes play a fundamental role in impacting student performance during cognitive tasks. The current study discusses severity levels of frustration, anxiety, depression, and stress. The impact of frustration is the natural part of learning as well as the

engaging session (for references, see the literature review section). Such emotion is always found during comprehensive cognitive activities. Literature saturated with many qualitative findings focused on the statistical association between student performance and frustration. However, the study has not been evidenced a comprehensive approach to solve the challenges produced by frustration during cognitive tasks.

We need to analyze the performance of excellent, average, and at-risk students while mathematically modeling the relationship between institution-related attributes, students' emotional factors, and family-related attributes. Also, the teacher can help frustrated students' through collaborative exercises, group activities, and group assignments [123]. It will help students easily share their confusion and problems with group members to overcome their frustrations in a comprehensive learning environment. An individual can learn better in offline mode with face-to-face interaction as compared to online interaction [124]. Additionally, the COVID-19 outbreak has accelerated the influence of negative emotions on students' performance. COVID-19 has created a more critical situation for students' learning and adjusted them to the online environment with fewer resources. Thus, we are in dire need to evaluate the academic development of students while statistically associating the aforementioned factors and mathematically modeling the proposed relationship to prepare for the critical situation [125].

4.2.3. Gender Description. In the literature review section, the study has shown that earlier studies statistically associated students' performance with emotional attributes and gender description. Students perform differently while considering aging and gender [126]. Both emotion and gender need to evaluate differently during cognitive activities. Literature studies are evidenced with many contributions on gender differences. They show that different gender individuals perform differently during cognitive activities, solving assignments, attempting quizzes, and examinations. Earlier studies depict that gender difference is an independent biological factor whose magnitude is sometimes dependent on other factors such as cultures, socioeconomic condition, language, age, etc. Gender differences play a crucial role in influencing mental abilities and cognitive processing in mathematical tasks, physics, research, reading, and writing. These issues create a big gap between male and female individuals, referred to as natural and biological differences.

4.2.4. Institution-Related Factors. Different institutional factors are directly or indirectly involved in influencing students' performance. These factors include but are not limited to instructor teaching methodology, interaction with a student advisor, extracurricular activities in the institution, student complaint platform, the distance between the institution and students' residence, transport facility, and the behaviors of the friends. These all factors have merits and demerits for student performance. The literature studies of

psychology and data analysis have enormous contributions to student performance analysis; however, insignificant contributions have been reported in the form of algorithms and mathematical models in students' performance prediction.

4.3. Q3: What Is the Intensity of Research Findings in the field of Student Performance Prediction Systems Optimization? Literature reported many challenges because the students' performance prediction overlaps psychology, data analysis, and mathematical and algorithmic contributions. The intensity of publications in the student performance prediction area is reported below.

4.3.1. Intensity of Psychological Findings. As discussed earlier, we can find many psychological research contributions in the field of student performance analysis, which show that emotional attributes always affect students' performance during cognitive activities. So, to provide an efficient solution for student performance prediction, the study must need to evaluate the psychological findings that directly or indirectly focus on student performance evaluation.

4.3.2. Intensity of Data Analysis Findings. Data analysis contribution provides a quantitative measurement for student performance prediction. Such research findings pave the way for an accurate mathematical model to better contribute to the performance prediction area of research.

4.3.3. Intensity of Students' Performance Prediction Systems. The literature is also saturated with student performance prediction techniques focusing on students' performance prediction in critical cognitive tasks; nevertheless, these findings are not synchronized and linked toward a significant student performance model. So, the main objective of this review paper is to provide an effective platform for future researchers in student performance prediction. It will pave the way for an effective system to predict at-risk students and excellent student performances, which ultimately provides us with the opportunity to enhance their skills and performance.

4.4. Q4: Are the Findings of Psychological Studies, Data Mining, and Contribution in Algorithms Are Synchronized with Each Other for the Viability of Pilot Project? The intensity of publications contributing to student performance prediction is quite good, but these contributions are not synchronized with each other to mathematically model emotional, family, and institution-related attributes. One of the main objectives of the current review is to highlight the lack of coordination and synchronization of the literature from different research fields. This review would allow future readers of deep learning to collaborate with other research fields.

4.5. Q5: How Synchronization and Coordination of Prior Psychological, Data Mining, and Algorithmic Findings Contribute to the Effective Educational System via Student Performance Prediction Algorithm. Psychological literature produces both qualitative and quantitative findings in students' performance prediction; nevertheless, the data analysis field highlights the association among students' factors, i.e., emotional, family, and institutional attributes. If these findings are linked with the objective of qualitative data repositories and algorithms, then, we can move toward an efficient student performance prediction system. The psychological work produces accurate students' emotional data focusing on their performance. On the other hand, the data analysis field makes the meaningful statistical association and correlation information. The data analysis field of research provides a couple of tests to find the correlation between student emotional attributes and their performance, i.e., Pearson correlation and regression. These tests verify the correlation among different factors.

We are in dire need to have the abovementioned psychological and data analysis findings to propose a comprehensive algorithm. Every part of the student performance prediction area of research is interlinked. The psychological result verifies the emotional change during the evaluations of the frustration, severity, anxiety, and stress. Second, the data analysis findings associate the student attributes. Third, the student performance prediction algorithm mathematically model the statistical association among the student influencing factors and their performance outcome.

4.6. Specific Keywords-Wise Publications. This section intensively discusses the specific keyword-wise research output focusing on students' performance, emotional factors, and prediction algorithms. The list of keywords is illustrated in Table 1. The study collected articles based on these keywords for further technical assessment. The specific domain for the technical evaluation includes but is not limited to new methods, modifications in prior work, data analysis, psychological findings, application analysis, review work, and comparison. The self-explanatory Table 1 illustrated the intensity of publications in the domain above using the list of keywords.

4.7. Yearly Publications. Literature studies deliver thousand of research findings that directly or indirectly contribute to students' performance analysis and prediction. As illustrated in Figure 2, 37 published articles were evaluated (1990 to 1994). About 110 articles mainly focus on student performance and students' study-related factors assessment. They have evaluated those factors that affect students' performance during cognitive activities (1995 to 1999). From 2000 to 2004, the study included 144 articles on students' performance and emotional attributes. The number of featured articles increases with time. From 2005 to 2009, we have assessed 310 articles that contributed to performance prediction.

Furthermore, we have collected 557 research studies that mainly focused on prediction algorithms. The researchers

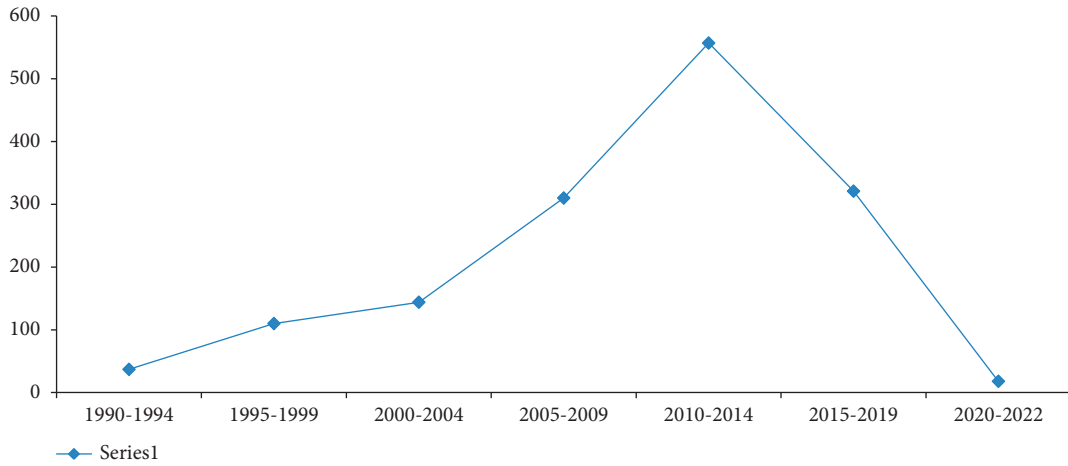


FIGURE 2: Yearly research contributions.

TABLE 2: Research domain-wise keyword searching results and evaluation.

	New methods	Modification	Data analysis	Psychological findings	Comparison	Analysis of application	Review work	Total number of publications
IEEE	15	23	68	65	24	10	9	214
ACM	14	16	65	70	13	14	9	201
Springer	11	17	87	32	4	7	8	166
MDPI	18	12	48	61	3	6	13	161
Hindawi	16	22	72	45	17	12	18	202
Elsevier	22	15	36	70	4	2	15	164
Wiley	8	9	66	32	14	56	17	202
Others	5	3	37	87	13	39	3	187

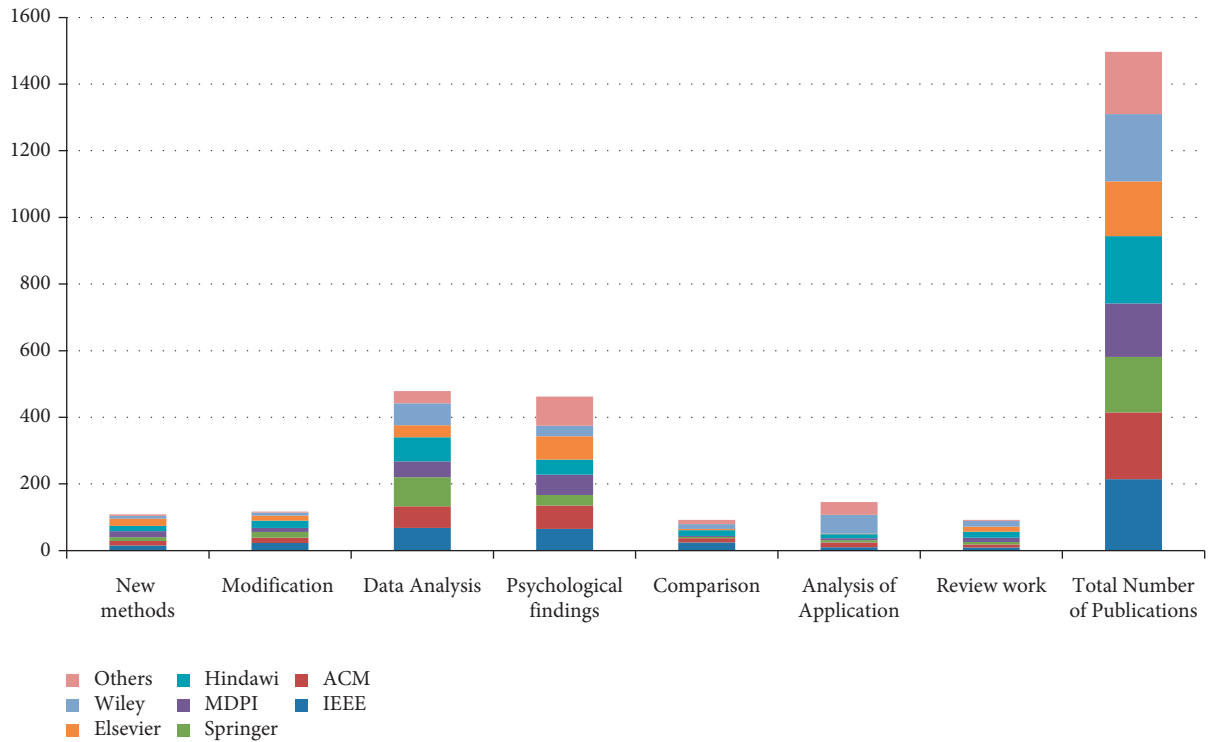


FIGURE 3: Domain-wise and publishers-wise outcomes.

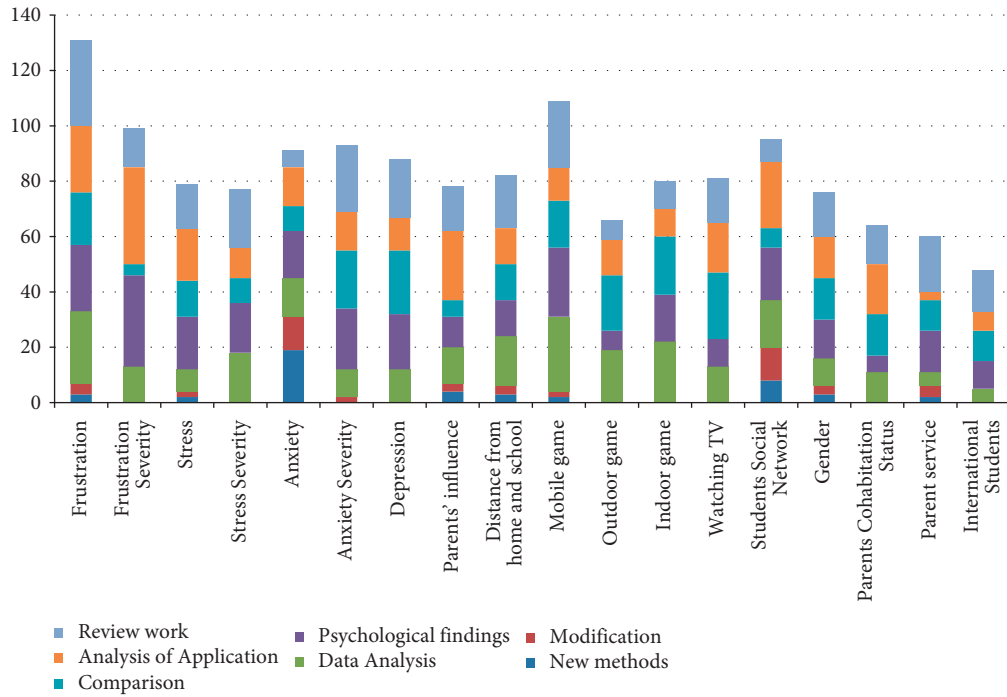


FIGURE 4: Factors-wise and domain-wise outcomes.

TABLE 3: Research domain-wise and factors-wise evaluation.

Attributes	New methods	Modification	Data analysis	Psychological findings	Comparison	Analysis of application	Review work	Total number of publications
Frustration	3	4	26	24	19	24	31	131
Frustration severity	0	0	13	33	4	35	14	99
Stress	2	2	8	19	13	19	16	79
Stress severity	0	0	18	18	9	11	21	77
Anxiety	19	12	14	17	9	14	6	91
Anxiety severity	0	2	10	22	21	14	24	93
Depression	0	0	12	20	23	12	21	88
Parents' influence	4	3	13	11	6	25	16	78
Distance from home and school	3	3	18	13	13	13	19	82
Mobile game	2	2	27	25	17	12	24	109
Outdoor game	0	0	19	7	20	13	7	66
Indoor game	0	0	22	17	21	10	10	80
Watching TV	0	0	13	10	24	18	16	81
Students social network	8	12	17	19	7	24	8	95
Gender	3	3	10	14	15	15	16	76
Parents cohabitation status	0	0	11	6	15	18	14	64
Parent service	2	4	5	15	11	3	20	60
International students	0	0	5	10	11	7	15	48

delivered a considerable amount articles from 2010 to 2014. We have found a slight decrease in analysis and psychological studies production until 2019; however, an increase was observed in algorithmic work from 2015 to 2019. Thus, 321 studies were considered during this segment of time. Eventually, the study reviewed 18 articles on students' performance prediction algorithms published from 2020 to

2022. We have collected 1497 articles during the review process of the current study.

4.8. Domain-Wise Evaluation. The review paper carries out a perceptive analysis and literature-based survey to draw out an inclusive representation of the famous publishers who

TABLE 4: Intensity of various domain contributions.

Research outcomes	New methods	Modification	Data analysis	Psychological findings	Comparison	Analysis of application	Review work	Total number of publications
SRI		3			2			5
PSP-PRP	5						4	9
AS-EDM			3					3
MAR-LD			4			6		10
RFA		6		4				10
MAR-KD			7					7
AF-TM							1	1
PSP-GPA	6			1			1	8
PSP-MC			5	5				10
EDU-DMCR		4	5					9
UPSP-EDU			6					6
PSP-OCS		7						7
PC-SD			5		5			10
EDU-DMA					3			3
EDU-DMLA						6		6
SEDU-DM			3					3
EDU-DMA			5					5
SRCT						7	3	10
PSBS	1	3					2	6
EITE		5						5
ICAP						7		7
TEO					5			5
LAD		5						5
IEDU-PM			6					6
SAFP		5				2	3	10
SPIE				9				9
PSS-TEDC			7				2	9
DMA-SD		4			2			6
EDU-DMPC		4					3	7
PSP-CA			6					6
PSS-CF	2				3		2	7
DMM-SC		1				4		5
PMSA				8				8
QACL	1		2	3			1	7
HSE				7				7
PPM-AS	2	3						5
RAE			8					8
ICASP			2		7			9
PSP-TDF	1	3				2		6
SSC					4			4
EDPLC					5			5
KPS-EDU		5				3		5
PSP-NBT	1	2						6
SMA		2					5	7
SMCS	1	1						2
PSP-LMS	5	4				5		14
TLs			2			5		7
ARICM			2					2
PSP-ALA		5						5
SARL		3					5	8
EDM		2				2		4
TRI-PAP			3		2			5
PSP-PA		2						2
AWP							3	3
TSBCP		3						3
AUD			5					5
CPU					3		2	5
DDC	1					3		4
EDMD-APS			2					2

TABLE 4: Continued.

Research outcomes	New methods	Modification	Data analysis	Psychological findings	Comparison	Analysis of application	Review work	Total number of publications
MA-TE					2			2
PRD-SF-GP				2				2
SE-UT					1		6	7
SE-DRVPU				3				3
PRD-DFA	3				3			6
AANL-PISA		4						4
INT-INF							4	4
PRD-GR				4		2		6
ESEG-AP		2			1	5		8
SAG-EDM	2							2
MSD-PRD		1					3	4
PRI-SRMA	3							3
SRS-AL		3		4		4		11
SET-IFP				6				6
SRL-HYPM							2	2
SAP-DM	4			4				8
LAS-TEL		1					3	4
DMKMS		2			6			8
DSS-LE						5		5
FGCAC				2				2
SAPM	2							2
DM-PSP						2		2
OPCA	1							1
HESSP-PP		3						3
IOMC			2					2
DM-CRTL						2		2
DM-ED				5				5
FGSK-SP	2							2
EDM-ARW			3			6		9
GP-SSM	1							1
FENTP				3				3
TE-LMSF		3				5		8
RGTE							3	3
DOF-DTT	1							1
P-CSI				3				3
DTDM	1	1						2
PSP-SDMA	5	4				2		11
SAS					6			6
ODF-AFQP			2					2
PSP-OLDF		1						1
EDM-S		3					5	8
EDM-RSA		1				2		3
ASP-DCBC			4		2			6
CSP-LCV		2						2
EEDM-IPC	2							2
PSP-C				3	3			6
PSP-DMT	2					8		10
WUGC						2		2
SEDM-PSP							4	4
LAEDM-CC						2		2
PSP-EDT			3			5		8
TQSA-ES	4					4		8
PMTF							3	3
DFUS				3				3
SPP-CS	1					2		3
MED-CS		4						4
IAPP	1				4			5
MM-SN						7		7
TQ-CS		5		3				8

TABLE 4: Continued.

Research outcomes	New methods	Modification	Data analysis	Psychological findings	Comparison	Analysis of application	Review work	Total number of publications
MA-FTE	2		2					4
FGAM						7		7
MRF-CA	2		3					5
GSM	3			2			3	3
OCM							3	3
LRMP				4				4
IDK		7			4			11
EDM				2				2
ILA-EDM							5	5
RPP	1		4			2		7
SP- RBFNN&PCA			4					4
SP-MLR&PCA	4					5		9
WTM			1		6			7
SCS		4						4
LF-PP				2				2
SA						5		5
SET	4				4			8
WBLC		1		2				3
MRC						3		3
MBA-GL							2	2
SPM			4					4
S-GPA				5		1		6
PFDM							6	6
DMT-SN		3						3
PPS-COVID-19				4				4
ATI-F		2						2
NA-FD- COVID-19					4		3	7
COVID-19-AS	2							2
PI-COVID-19			2					2
SD-COVID-19					2			2
Edu-COVID-19			2					2
NCAS-COVID- 19							2	2
Imp-COVID-19			2	3				5
SS-PPP-DM	1					8		9
A-EDM-TD			3					3
RPSP-DMT							3	3
SPP-CL						6		6
ER-KCP		3						3
SDP				11				11
EDP-DM				4				4
PAP-SH	7							7
HMRS		2						2
IGR-PSP		3						3
PSPP-ML						2		2
ECE-RL	3		3	2			2	10
SML-OC							2	2
Inf-COVID-19					3			3
PEEP-COVID- 19			2			5		7
ATI-F		4					5	9
CCI-OC					4			4
ETES-COVID- 19		3		2				5
TFL-SF						4		4
OC-BL		6						6
NP-PSP						5		5

TABLE 4: Continued.

Research outcomes	New methods	Modification	Data analysis	Psychological findings	Comparison	Analysis of application	Review work	Total number of publications
SPP-BL			2					2
RSNL				6				6
DN-CS	1		3					4
PR-MS				7				7
DSF-HB	1		2	3			6	12
VFP-C				7				7
EAK-P	2	3						5
SP-EG-MM			8					8
LMS-CAP			2		7			9
FDG	2							2
BFE				3				3
AD-CS	2	3						5
SP-ALA			8					8
TVL-CA			2		7			9
EAG-CSC				4				4
ML-CSC			2					2
SP-DM-LAT	4		5					9
S-GC			4					4
MR-PCQ				3				3
MPA-M	1					2		3
MCA-E		4					3	7
T-PR				3				3
NS-SE						5		5
ARFE			3					3
CSMA				8				8
NT-PPCS	3							3
MSG-IC					5			5
GD-ATC		3						3
GD-AT-SCI			3					3
LS-ESP-R			4					4
GD-SE			2					2
GD-AT-IT				4				4
GD-RC		3				2		5
GD-LTS			2					2
SG-TM-CAP	2					4		6
GDSL			4					4
GD-MS-SL		3					5	8
GD-MR				5				5
DSS-CP	2	6				2		10
GD-TET			4					4
GD-HSS			2		3			5
TP-MA				4			3	7
GD-NCS	1		2					3
GD-SP-EC				4				4
SSG		4						4
GD-DSS			3			5		8
ETP-SSA				5				5
GES-E			3					3
PSD		2						2
IQ-PAP					4			4
TSI-SSC							3	3
BFP-MA		3					4	7
ESF-SS				2				2
FPP-AUS			5	6				11
ACA			3				6	9
AAGT	1	3		5				9
SLC-A		2					3	5
RHAS				8				8
PSO-LPS	1	2						3

TABLE 5: Abbreviation and acronym.

Abbreviation	Acronym
SRI	The dimensionality of student ratings of instruction: what we know and what we do not
PSP-PRP	Predicting student performance on post-requisite skills using prerequisite
AS-EDM	An approachable analytical study on big educational data mining
MAR-LD	Mining association rules between sets of items in large databases
RFA	Clarify of the random forest algorithm in an educational field
MAR-KD	Knowledge discovery from academic data using association rule mining
AF-TM	How automated feedback through text mining changes plagiaristic behavior in online assignments
PSP-GPA	Predicting students final GPA using decision trees
PSP-MC	Analyzing students performance using multicriteria classification
EDU-DMCR	Data mining in educational technology classroom research
UPSP-EDU	Analyzing undergraduate students' performance using educational data mining
PSP-OCS	Student performance prediction and optimal course selection
PC-SD	Probabilistic classifiers and statistical dependency
EDU-DMA	Educational data mining: an advance for intelligent systems in education
EDU-DMLA	Educational data mining and learning analytics
SEDU-DM	The state of educational data mining in 2009
EDU-DMA	Educational data mining applications and tasks
SRCT	Student ratings of college teaching
PSBS	Predicting drop-out from social behavior of students
EITE	Ensemble learning for estimating individualized treatment effects in student success studies
ICAP	Identifying the comparative academic performance of secondary schools
TEO	Taxonomy of educational objectives
LAD	The design, development, and implementation of student-facing learning analytics dashboards
IEDU-PM	Clustering for improving educational process mining
SAFP	Determining students' academic failure profile founded on data mining methods
SPIE	Student perceptions and instructional evaluations
PSS-TEDU	Predicting student success using data generated in traditional educational environments
DMA-SD	Data mining application on students' data
EDU-DMPC	Educational data mining for prediction and classification of engineering students achievement
PSP-CA	A comparative analysis of techniques for predicting student performance
PSS-CF	Predicting students success in courses via collaborative filtering
DMM-SC	Data mining models for student careers
PMSA	Blending measures of programming and social behavior into predictive models of students achievement in early computing courses
QACL	Quantitative approach to collaborative learning
HSE	Will teachers receive higher student evaluations by giving higher grades and less course work?
PPM-AS	Student performance prediction model for early-identification of at-risk students in traditional classroom settings
RAE	Regression analysis by example
ICASP	Mining the impact of course assignments on student performance
PSP-TDF	Predicting student performance in an ITS using task-driven features
SSC	Soft subspace clustering of categorical data with probabilistic distance
EDPLC	Early detection prediction of learning outcomes in online short-courses via learning behaviors
KPS-EDU	Tracking knowledge proficiency of students with educational priors
PSP-NBT	Exploration of classification using NB tree for predicting students' performance
SMA	Student modeling approaches: a literature review for the last decade
SMCS	An ontological approach for semantic modeling of curriculum and syllabus in higher education
PSP-LMS	Predicting student performance from LMS data
TLR	Organizing knowledge syntheses: a taxonomy of literature reviews
ARICM	Analysis of academic results for informatics course improvement using association rule mining
PSP-ALA	Predicting student performance using advanced learning analytics
SARL	Seeding the survey and analysis of research literature with text mining
EDM	A systematic review of educational data mining
TRI-PAP	Do the timeliness, regularity, and intensity of online work habits predict academic performance?
PSP-PA	Predicting student performance using personalized analytics
AWP	Automated analysis of aspects of written argumentation
TSBCP	Predicting performance form test scores using back propagation and counter propagation
AUD	The text mining handbook: advanced approaches in analyzing unstructured data, cambridge
CPU	Cell phone usage and academic performance
DDC	Learning analytics: drives, developments and challenges
EDMD-APS	Educational data mining discovery standards of academic performance by students

TABLE 5: Continued.

Abbreviation	Acronym
EDM-PAAP	Educational data mining: predictive analysis of academic performance
HGS-AFS	Do high grading standards affect student performance?
RHS	Retrieving hierarchical syllabus items for exam question analysis
SE-TE	Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes?
DAS-AARM	Drawbacks and solutions of applying association rule mining in learning management systems
MPAP	Model prediction of academic performance for first year students
EPMSS	Evaluating predictive models of student success: closing the methodological gap
LA	Learning analytics should not promote one size fits all
DLA	Detecting learning strategies with analytics: links with self-reported measures and academic performance
SGP-NN	Explaining student grades predicted by a neural network
PAP	Predicting academic performance
MTQ	Measuring teaching quality in higher education
AD-SLS	Towards automatically detecting whether student learning is shallow
CMPL	An application of classification models to predict learner progression in tertiary education
USWT	Utilizing semantic web technologies and data mining techniques to analyze students learning and predict final performance
LAP	A model to predict low academic performance at a specific enrollment using data mining
PSP	Predicting students performance in educational data mining
NSP-KDHED	A new student performance analysing system using knowledge discovery in higher educational databases.
MLM	Comparison of machine learning methods for intelligent tutoring systems
ID-CS	Individual differences related to college students' course performance in calculus
SAP	Student academic performance prediction by using a decision tree algorithm.
PP-PSP	Performance prediction based on particle swarm optimization
PSP-M	Poverty and student performance in Malaysia
PA-PS	Physical activity is not related to performance at school
PF	The power of feedback, review of educational research
IDF-SAP	Identifying key factors of student academic performance by subgroup discovery
SC-NF	Student classification for academic performance prediction using neuro fuzzy in a conventional classroom
OEP-TRF	Online education performance prediction via time-related features
PCS	Programming content semantics: an evaluation of visual analytics approach
SVA-PC	Semantic visual analytics for today's programming courses
PSL	A systematic review of studies on predicting student learning outcomes using analytics
SAP-EDC	Predicting student academic performance in an engineering dynamics course: a comparison of four types of predictive mathematical models
PRD-AP	Predicting student's academic performance: comparing artificial neural network, decision tree, and linear regression
SSP-CL	Analyzing student spatial deployment in a computer laboratory
QE-ELC	Quality enhancement for e-learning courses: the role of student feedback
GRP-OEWB	Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique
SP-DMC	Student performance prediction by using data mining classification algorithms
PPRD-DT	Performance prediction of engineering students using decision trees
SUR-MSR	A survey and taxonomy of approaches for mining software repositories in the context of software evolution
SPRD-ARMBA	A review and performance prediction of students' using an association rule mining based approach
EXP-HPF	Exploring the high potential factors that affects students' academic performance
IPT-SP	Analysing the impact of poor teaching on student performance
DM-ETSP	Data mining based analysis to explore the effect of teaching on student performance
SPP-DL	Gritnet: student performance prediction with deep learning
DM-E	Data mining and education
PSM-HOU	Predicting students marks in hellenic open university
PSP-ML	Predicting postgraduate students' performance using machine learning techniques
PA-EDM	Review on prediction algorithms in educational data mining
LS-PRDE	Literature survey on student's performance prediction in education using data mining techniques
PRD-AP	Predicting student academic performance
HSC-SA	Online self-paced high-school class size and student achievement
PRI-MPP	Predictor relative importance and matching regression parameters
SE-OES	Finding similar exercises in online education systems
FCD-EP	Fuzzy cognitive diagnosis for modeling examine performance
EB-PSP	An ensemble-based semi-supervised approach for predicting students' performance
MSM-ENB	Measuring the (dis-) similarity between expert and novice behaviors as serious games analytics

TABLE 5: Continued.

Abbreviation	Acronym
M-KME	Mining for topics to suggest knowledge model extensions
EPRD-BL	Applying learning analytics for the early prediction of students' academic performance in blended learning
MA-TE	Whose feedback? A multilevel analysis of student completion of end-of-term teaching evaluations
PRD-SF-GP	Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data
SE-UT	Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness
SE-DRVPU	Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness
PRD-DFA	Predicting student outcomes using discriminant function analysis
AANL-PISA	An overview of using academic analytics to predict and improve students' achievement: a proposed proactive intelligent intervention
INT-INF	Constructing interpretive inferences about literary text: the role of domain-specific knowledge
PRD-GR	Predicting grades
ESEG-AP	Early segmentation of students according to their academic performance: a predictive modeling approach
SAG-EDM	A framework for smart academic guidance using educational data mining
MSD-PRD	Mining students' data for prediction performance
PRI-SRMA	Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement
SRS-AL	A semantic recommender system for adaptive learning
SET-IFP	Students evaluating teachers: exploring the importance of faculty reaction to feedback on teaching
SRL-HYPM	Self-regulated learning with hypermedia: the role of prior domain knowledge
SAP-DM	Modeling and predicting students' academic performance using data mining techniques
LAS-TEL	Lexical analysis of syllabi in the area of technology enhanced learning
DMKMS	Student data mining solution-knowledge management system
DSS-LE	Decoding student satisfaction: how to manage and improve laboratory experience
FGCAC	Student ability best predicts final grade in a college algebra course
SAPM	Student academic performance monitoring and evaluation
DM-PSP	Data mining approach for predicting student performance
OPCA	Optimizing partial credit algorithms
HESSP-PP	Is alcohol affecting higher education students' performance: searching and predicting pattern
IOMC	Towards the integration of multiple classifier pertaining to the student's performance prediction
DM-CRTL	A data mining view on classroom teaching language
DM-ED	Application of data mining in educational databases for predicting academic trends and patterns
FGSK-SP	Using fine-grained skill models to fit student performance
EDM-ARW	Educational data mining: a survey and a data mining-based analysis of recent works
GP-SSM	Grade prediction with course and student specific models
FENTP	Feature extraction for next-term prediction of poor student performance
TE-LMSF	Teaching evaluation using data mining on moodle LMS forum
RGTE	The role of gender in students' ratings of teaching quality in computer science and environmental engineering
DOF-DTT	Drop out feature of student data for academic performance using decision tree techniques
P-CSI	Programming: predicting student success early in CSI
DTDM	Decision trees and decision-making
PSP-SDMA	Predicting student performance: a statistical and data mining approach
SAS	A sentiment analysis system to improve teaching and learning
ODF-AFQP	Ontology driven framework for assessing the syllabus fairness of a question paper
PSP-OLDF	Predicting students' final performance from participation in on-line discussion forums
EDM-S	Educational data mining: a survey from 1995 to 2005
EDM-RSA	Educational data mining: a review of the state of the art
ASP-DCBC	Analyzing student performance using sparse data of core bachelor courses
CSP-LCV	Centralized student performance prediction in large courses based on low-cost variables in an institutional context
EEDM-IPC	Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses
PSP-C	Prediction of students' academic performance using clustering
PSP-DMT	A review on predicting students' performance using data mining techniques
WUGC	Web-based undergraduate chemistry problem-solving: the interplay of task performance, domain knowledge and web-searching strategies
SEDM-PSP	A survey on various aspects of education data mining in predicting student performance
LAEDM-CC	Learning analytics and educational data mining: towards communication and collaboration
PSP-EDT	Predictive modeling of students performance through the enhanced decision tree
TQSA-ES	What is the relationship between teacher quality and student achievement? An expletory study
PMTP	A predictive model for standardized test performance in Michigan schools
DFUS	Determination of factors influencing the achievement of the first-year university students

TABLE 5: Continued.

Abbreviation	Acronym
SPP-CS	Next-terms student performance prediction: a case study
MED-CS	Mining educational data to improve students' performance: a case study
IAPP	Improving academic performance prediction by dealing with class imbalance
MM-SN	Proposing stochastic probability-based math model and algorithms utilizing social networking and academic data
TQ-CS	Teaching quality matters in higher education: a case study
MA-FTE	Meta-analysis of faculty's teaching effectiveness: student evaluation of teaching ratings and student learning
FGAM	Analysis of the impact of action order on future performance: the fine-grain action model
MRF-CA	Map-reduce framework based cluster architecture for academic students' performance prediction
GSM	Google Scholar coverage of a multidisciplinary field
OCM	The opportunity count model: a flexible approach to modeling student performance
LRMP	Predicting students' performance in final examination using linear regression and multilayer perceptron
IDK	Fast searching for information on the internet to use in a learning context: the impact of domain knowledge
EDM	Educational data mining acceptance among undergraduate students
ILA-EDM	Participation-based student final performance prediction model through interpretable genetic programming: integrating learning analytics, educational data mining and theory
RPP	Improving retention performance prediction with prerequisite skill features
SP-RBFNN&PCA	Predicting honors student performance using RBFNN and PCA method
SP-MLR&PCA	Predicting students' academic performance using multiple linear regression and principal component analysis
WTM	Web-based collaborative writing in L2 contexts: methodological insights from text mining
SCS	Chinese undergraduates' perceptions of teaching quality and the effects on approaches to studying and course satisfaction
LF-PP	Can online discussion participation predict group project performance? Investigating the roles of linguistic features and participation patterns
SA	Improving early prediction of academic failure using sentiment analysis on self-evaluated comments
SET	The use and misuse of student evaluations of teaching
WBLC	A multivariate approach to predicting student outcomes in web-enabled blended learning courses
MRC	Mendeley: creating communities of scholarly inquiry through research collaboration
MBA-GL	A model-based approach to predicting graduate-level performance using indicators of undergraduate-level performance
SPM	Students performance modeling based on behavior pattern
S-GPA	Predicting students' GPA and developing intervention strategies based on self-regulatory learning behaviors
PFDM	Towards parameter-free data mining: mining 'educational data with yacaree
DMT-SN	A survey of data mining techniques for social network analysis
PPS-COVID19	New realities for polish primary school informatics education affected by COVID-19
ATI-F	Affect-targeted interviews for understanding student frustration
NA-FD-COVID19	Unhappy or unsatisfied: distinguishing the role of negative affect and need frustration in depressive symptoms over the academic year and during the COVID-19 pandemic
COVID19-AS	COVID-19 disruption on college students: academic and socioemotional implications
PI-COVID19	The psychological impact of COVID-19 on the mental health of the general population
SD-COVID19	Social distancing in covid-19: what are the mental health implications?
Edu-COVID19	Education and the COVID-19 pandemic
NCAS-COVID19	Negative emotions, cognitive load, acceptance, and self-perceived learning outcome in emergency remote education during COVID-19
Imp-COVID19	The impact of COVID-19 on education insights from education at a glance 2020
SS-PPP-DM	Study on student performance estimation, student progress analysis, and student potential prediction based on data mining
A-EDM-TD	Application of educational data mining approach for student academic performance prediction using progressive temporal data
RPSP-DMT	A review on predicting students' performance using data mining techniques
SPP-CL	Student performance analysis and prediction in classroom learning: a review of educational data mining studies
ER-KCP	Exercise recommendation based on knowledge concept prediction
SDP	Student dropout prediction
EDP-DM	Early dropout prediction using data mining: a case study with high school students
PAP-SH	Predicting academic performance by considering student heterogeneity
HMRS	Helping university students to choose elective courses by using a hybrid multicriteria recommendation system with genetic optimization
IGR-PSP	Inductive Gaussian representation of user-specific information for personalized stress-level prediction
PSPP-ML	Pre-course student performance prediction with multi-instance multi-label learning
ECE-RL	What students want? Experiences, challenges, and engagement during emergency remote learning amidst COVID-19 crisis

TABLE 5: Continued.

Abbreviation	Acronym
SML-OC	A survey of machine learning approaches for student dropout prediction in online courses
Inf-COVID19	Covid-19 and student performance, equity, and us education policy: lessons from pre-pandemic research to inform relief, recovery, and rebuilding
PEEP-COVID19	COVID19 and student performance equity, and us education Policy: Lessons from pre-pandemic research to inform relief, recovery, and rebuilding.
ATI-F	“Affect-targeted interviews for understanding student frustration”, in international conference on artificial intelligence in education
CCI-OC	Common challenges for instructors in large online course: strategies to mitigate student and instructor frustration
ETES-COVID19	Effective teaching and examination strategies for undergraduate learning during COVID-19 school restrictions
TFL-SF	Teacher feedback literacy and its interplay with student feedback literacy
OC-BL	Challenges in the online component of blended learning: a systematic review
NP-PSP	Feature extraction for next-term prediction of poor student performance
SPP-BL	Student performance prediction based on blended learning
RSNL	Robust student network learning
DN-CS	Deep network for the iterative estimations of students’ cognitive skills
PR-MS	Parents’ role in the academic motivation of students with gifts and talents
DSF-HB	Detecting student frustration based on handwriting behavior
VFP-C	The validity of a frustration paradigm to assess the effect of frustration on cognitive control in school-age children
EAK-P	Ekt: exercise-aware knowledge tracing for student performance prediction
SP-EG-MM	Predicting student performance in an educational game using a hidden Markov model
LMS-CAP	Massive lms log data analysis for the early prediction of course-agnostic student performance
FDG	Frustration drives me to grow
BFE	Between frustration and education: transitioning students’ stress and coping through the lens of semiotic cultural psychology
AD-CS	Automatic discovery of cognitive skills to improve the prediction of student learning
SP-ALA	Predicting student performance using advanced learning analytics
TVL-CA	Time-varying learning and content analytics via sparse factor analysis
EAG-CSC	Emotions, age, and gender based cognitive skills calculations
ML-CSC	Machine learning based cognitive skills calculations for different emotional conditions
SP-DM-LAT	Predicting student performance using data mining and learning analytics techniques: a systematic literature review
S-GC	Should I grade or should I comment: links among feedback, emotions, and performance
MR-PCQ	Modeling the relationship between students’ prior knowledge, causal reasoning processes, and quality of causal maps
MPA-M	A multilayer prediction approach for the student cognitive skills measurement
MCA-E	A meta-cognitive architecture for planning in uncertain environments
T-PR	The influence of teacher and peer relationships on students
NS-SE	National Society for the Study of Education
ARFE	Automatically recognizing facial expression: predicting engagement and frustration
CSMA	A biologically inspired cognitive skills measurement approach
NT-PPCS	A novel technique for the evaluation of posterior probabilities of student cognitive skills
MSG-IC	Medical student gender and issues of confidence
GD-ATC	Gender differences in student attitudes toward computers
GD-AT-SCI	Gender differences in student attitudes toward science: a meta-analysis of the literature from 1970 to 1991
LS-ESP-R	A longitudinal study of engineering student performance and retention III. Gender differences in student performance and attitudes
GD-SE	Gender differences in student ethics: Are females really more ethical? Gender differences in teacher-student interactions in science classrooms
GD-AT-IT	Gender differences in attitudes towards information technology among Malaysian student teachers: a case study at University Putra Malaysia
GD-RC	Gender differences in the response to competition
GD-LTS	Gender differences in the learning and teaching of surgery: a literature review
SG-TM-CAP	Student gender and teaching methods as sources of variability in children’s computational arithmetic performance
GDSL	Gender difference and student learning
GD-MS-SL	Gender difference in student motivation and self-regulation in science learning: a multigroup structural equation modeling analysis
GD-MR	Gender differences in the influence of faculty-student mentoring relationships on satisfaction with college among African-Americans
DSS-CP	Differences of students’ satisfaction with college professors: the impact of student gender on satisfaction
GD-TET	Gender differences in teachers’ perceptions of students’ temperament, educational competence, and teachability
GD-HSS	Gender differences in factors affecting academic performance of high school students
TP-MA	Influence of elementary student gender on teachers’ perceptions of mathematics achievement

TABLE 5: Continued.

Abbreviation	Acronym
GD-NCS	Gender differences in alcohol-related non-consensual sex, cross-sectional analysis of a student population
GD-SP-EC	Gender differences in students' and parents' evaluative criteria when selecting a college
SSG	Social influences, school motivation, and gender differences: an application of the expectancy-value theory
GD-DSS	Gender differences in the dimensionality of social support
ETP-SSA	Early teacher perceptions and later student academic achievement
GES-E	Gender, ethnicity, and social cognitive factors predicting the academic achievement of students in engineering
PSD	Predicting students drop out: a case study
IQ-PAP	Self-discipline outdoes IQ in predicting academic performance of adolescents
TSI-SSC	Observations of effective teacher-student interactions in secondary school classrooms: predicting student achievement with the classrooms assessment scoring system-secondary
BFP-MA	Role of the big five personality traits in predicting college students' academic motivation and achievement
ESF-SS	Using emotional and social factors to predict student success
FPP-AUS	Who succeeds at university? Factors predicting academic performance in first-year Australian university students
ACA	Predicting academic achievement with cognitive ability
AAGT	Advancing achievement goal theory: using goal structures and goal orientations to predict students' motivation, cognition, and achievement
SLC-A	Short-term and long-term consequences of achievement goals: predicting interest and performance over time
RHAS	Role of hope in academic and sports achievement
PSO-LPS	Prediction of school outcomes based on early language production and socioeconomic factors

TABLE 6: Summary of potential research challenges and recommendation.

S.No	Research question	Remarks	Recommendations
1	What are the applications of student performance prediction systems?	<p>Prediction of at-risk students for special treatment and counseling sessions.</p> <p>If students cannot achieve an excellent academic score, then the performance prediction system assists students in observing the main reason behind the low performance.</p> <p>Advance students' academic achievements.</p> <p>Monitor students' behavior such as interaction and attitude towards teacher, seriousness, and unseriousness in the classroom</p> <p>They include but are not limited to family-related factors, emotional factors, gender description, and institution-related factors.</p> <p>Emotional factors, such as frustration, anxiety, stress, and depression.</p>	<p>Mathematically model emotional attributes, family issues, study schedules, and institutional attributes all together to develop a significant prediction system.</p> <p>If the prediction system considers a large number of influential factors, then the academic achievement of excellent can also be advanced.</p> <p>Modulates the relationship between behavior and students' performance</p>
2	What are the factors that can optimize student performance prediction?	<p>Quantize family factors, i.e., parents' positive and negative roles, including overexpectation of parents and positive involvement of parents in children's daily cognitive activities.</p> <p>Literature studies are evidenced with many contributions to gender differences. They show that different gender individuals perform differently during cognitive activities, solving assignments, attempting quizzes, and examinations studies.</p> <p>Different institutional factors directly or indirectly influence students' performance.</p>	<p>Initiate pilot projects with an assumption-based dataset. The assumptions should be based on earlier studies of psychology, data analysis, and data mining.</p> <p>Analyze the performance of at-risk students while mathematically modeling the association among students' emotional, family, and institution-related attributes.</p> <p>Perform factorization of gender because earlier studies depict that gender difference magnitude is sometimes dependent on other factors such as cultures, socioeconomic condition, language, age, etc.</p> <p>Explore instructor teaching methodology, interaction with a student advisor, extra curriculum activities in the institution, student complaint platform, the distance between the institution and students' residence, transport facility, and the behavior of the friends.</p>

TABLE 6: Continued.

S.No	Research question	Remarks	Recommendations
3	What is the intensity of research findings in the field of student performance prediction systems optimization?	Intensity of psychological findings Intensity of data analysis findings	These findings are not synchronized and linked toward a significant student performance prediction model. So, the main challenge is to provide an effective platform where future researchers can collaborate and synchronize the prior findings. Also, pilot projects based on the assumption-based dataset are highly recommended. Successful pilot project implementation will pave the way for quick optimization of existing systems.
4	Are the findings of psychological studies, data mining, and contribution in algorithms synchronized with each other for the viability of the pilot project?	Intensity of students' performance prediction systems The intensity of publications contributing to student performance prediction is quite good, but these contributions are not synchronized with each other. Every part of the student performance prediction area of research is interlinked.	Mathematically model emotional, family, and institution-related attributes.
5	How do synchronization and coordination of prior psychological, data mining, and algorithmic findings contribute to the effective educational system via student performance prediction algorithm?	The psychological result verifies the emotional change during the evaluations of the frustration, severity, anxiety, and stress. The data analysis findings associate the student attributes. The student performance prediction algorithm mathematical model the statistical association among the student influencing factors and their performance outcome.	If these findings are linked with the objective of qualitative data repositories and algorithms, then, we can move toward an efficient student performance prediction system.

publish the most highly cited research papers. It involves overall 1497 publications, which are selected after searching on Google Scholar. These studies were evaluated upon their relevant findings, such as new methods, modified approaches, statistical findings, and psychological results (for more information see Table 2 and Figure 3). The major part of this review was to assess the existing students' performance prediction approaches. So, the study analytically assessed new students' performance prediction measures, modifications in state-of-the-art techniques, and comparative analysis. Additionally, Figure 4 and Table 3 represent the detailed domain-wise and factors-wise analysis.

5. Potential Future Challenges

A large number of factors are involved in influencing students' performance; therefore, the prediction system needs to be optimized to consider the impacts of different human factors categories. Such factors categories include but are not limited to emotional attributes, study schedule, family attributes, and institutional attributes. Each category consists of multiple factors impacting students' performance, either negatively or positively. In the literature section, the study provides a detailed discussion of these factors.

5.1. Potential Pilot Projects Based on the Assumption-Based Dataset. The comprehensive synchronization between the earlier studies is still a black box, which increases systems'

dependency on a real-world dataset. The importance of a real-world dataset cannot be avoided; however, the data collection process is time-consuming and need a list of human resources. It delays the optimization of existing approaches, such as modeling students' emotional attributes. The data collection process could have various anomalies if the researcher does not follow the analysis of earlier studies. The earlier studies offer excellent opportunities to understand the effectiveness of emotional attributes for optimization. Therefore, pilot projects perform key roles in optimizing the existing students' performance prediction systems. They provide useful ideas during the data collection process. They also pave the way for an assumption-based dataset to prove the viability of novel ideas in students' performance prediction.

6. Additional Points of Earlier Studies

Data analysis findings explore the hidden patterns and statistical correlation between students' performance and influential factors. Such opportunities introduce new challenges for students' performance prediction systems, e.g., conditional probabilities, correlation, and inferencing. Also, data mining studies are evidenced with many findings in students' performance prediction area of research; nevertheless, they have different limitations, e.g., lack of in-depth investigation of students' performance based on selected study-related factors, limited scalabilities, limited dataset, and inadequate qualitative approach of data analysis and psychological studies.

Finally, the review shows that various prior students' performance prediction methods have been proposed in the last decade; however, meagre studies have highlighted the basic need for synchronization among the abovementioned field's contributions. Therefore, this review provides an exclusive picture of the future challenges in students' performance prediction (see Table 1 to 4 and Figure 2 to 4). On one hand, Table 4 depicts the intensity of various optimization techniques, review works, and new students' performance prediction methods. On the other hand, Table 5 represents the acronyms of the selected studies. Remarks and recommendations against each research question are given in the self-explanatory Table 6.

7. Conclusions

The proposed review highlights the potential research opportunities to optimize the students' performance prediction systems while exploring earlier contributions of different research fields, i.e., cognitive computing, data mining, data analysis, and psychology. The previous studies are still limited in synchronization between the existing contributions of various fields, which negatively impacted the mathematical modeling of emotional attributes. It increased the systems' dependencies on real-world datasets. Thus, to investigate the potential challenges thoroughly, the study is split into three sections.

1. The data mining discoveries, psychological findings, and data analysis results are examined.
2. The study performs a domain-wise investigation of the existing methods focusing on students' performance prediction, i.e., the domain includes new students' performance prediction techniques, modifications in existing techniques, and comparisons analysis.
3. Eventually, future direction and potential pilot project viability are highlighted.

Data Availability

The screening data are available from the corresponding author, upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to acknowledge Prince Sultan University and the EIAS:Data Science and Blockchain Laboratory for their valuable support. Also, the authors would like to acknowledge the support of Prince Sultan University for the article processing charges (APCs) of this publication.

References

- [1] I. Sumanasekera, J. Abd Hamid, A. Khatibi, and S. F. Azam, "Involvement and style of parents on student motivation towards student performance with the moderating effect of academic causal factors: development of a conceptual model," *Global Journal of Management and Business Research*, vol. 21, no. 1, pp. 10–24, 2021.
- [2] M. Z. Osburn, C. Stegman, L. D. Suitt, and G. Ritter, "Parents' perceptions of standardized testing: its relationship and effect on student achievement," *Journal of Educational Research & Policy Studies*, vol. 4, no. 1, pp. 75–95, 2004.
- [3] K. V. Hoover-Dempsey, A. C. Battiato, J. M. T. Walker, R. P. Reed, J. M. DeJong, and K. P. Jones, "Parental involvement in homework," *Educational Psychologist*, vol. 36, no. 3, pp. 195–209, 2001.
- [4] H. Huang and G. Liang, "Parental cultural capital and student school performance in mathematics and science across nations," *The Journal of Educational Research*, vol. 109, no. 3, pp. 286–295, 2016.
- [5] L. C. Taylor, I. D. Hinton, and M. N. Wilson, "Parental influences on academic performance in african-american students," *Journal of Child and Family Studies*, vol. 4, no. 3, pp. 293–302, 1995.
- [6] L. Wößmann, "Schooling resources, educational institutions and student performance: the international evidence," *Oxford Bulletin of Economics and Statistics*, vol. 65, no. 2, pp. 117–170, 2003.
- [7] P. Mutodi and H. Ngirande, "The impact of parental involvement on student performance: a case study of a south african secondary school," *Mediterranean Journal of Social Sciences*, vol. 5, no. 8, p. 279, 2014.
- [8] S. J. Cabus and R. J. Ariës, "What do parents teach their children? - the effects of parental involvement on student performance in Dutch compulsory education," *Educational Review*, vol. 69, no. 3, pp. 285–302, 2017.
- [9] A. Harris and J. Goodall, "Do parents know they matter? engaging all parents in learning," *Educational Research*, vol. 50, no. 3, pp. 277–289, 2008.
- [10] R. M. A. Khan, N. Iqbal, and S. Tasneem, "The influence of parents educational level on secondary school students academic achievements in district rajanpur," *Journal of Education and Practice*, vol. 6, no. 16, pp. 76–79, 2015.
- [11] B. Basnet, M. Jaiswal, B. Adhikari, and P. M. Shyangwa, "Depression among undergraduate medical students," *Kathmandu University Medical Journal (KUMJ)*, vol. 10, no. 3, pp. 56–59, 2012.
- [12] C. Saravanan and R. Wilks, "Medical students' experience of and reaction to stress: the role of depression and anxiety," *The Scientific World Journal*, vol. 2014, Article ID 737382, 8 pages, 2014.
- [13] T. Alvi, F. Assad, M. Ramzan, and F. A. Khan, "Depression, anxiety and their associated factors among medical students," *Journal of the College of Physicians and Surgeons--Pakistan JCPSP*, vol. 20, no. 2, pp. 122–126, 2010.
- [14] T. L. Schwenk, L. Davis, and L. A. Wimsatt, "Depression, stigma, and suicidal ideation in medical students," *JAMA*, vol. 304, no. 11, pp. 1181–1190, 2010.
- [15] W. M. Chernomas and C. Shapiro, "Stress, depression, and anxiety among undergraduate nursing students," *International Journal of Nursing Education Scholarship*, vol. 10, no. 1, pp. 255–266, 2013.

- [16] K. Shamsuddin, F. Fadzil, W. S. W. Ismail et al., "Correlates of depression, anxiety and stress among Malaysian university students," *Asian journal of psychiatry*, vol. 6, no. 4, pp. 318–323, 2013.
- [17] K. L. Jansen, R. Motley, and J. Hovey, "Anxiety, depression and students' religiosity," *Mental Health, Religion & Culture*, vol. 13, no. 3, pp. 267–271, 2010.
- [18] D. P. Moreira and A. R. F. Furegato, "Stress and depression among students of the last semester in two nursing courses," *Revista Latino-Americana de Enfermagem*, vol. 21, pp. 155–162, 2013.
- [19] E. B. Davies, R. Morriss, and C. Glazebrook, "Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis," *Journal of Medical Internet Research*, vol. 16, no. 5, p. e130, 2014.
- [20] L. M. Al-Qaisy, "The relation of depression and anxiety in academic achievement among group of university students," *International Journal of Psychology and Counselling*, vol. 3, no. 5, pp. 96–100, 2011.
- [21] M. S. B. Yusoff, A. F. Abdul Rahim, A. A. Baba, S. B. Ismail, and M. N. Mat Pa, "Prevalence and associated factors of stress, anxiety and depression among prospective medical students," *Asian journal of psychiatry*, vol. 6, no. 2, pp. 128–133, 2013.
- [22] L. Chen, L. Wang, X. H. Qiu et al., "Depression among Chinese university students: prevalence and socio-demographic correlates," *PLoS One*, vol. 8, no. 3, Article ID e58379, 2013.
- [23] M. A. Moreno, L. A. Jelenchick, K. G. Egan et al., "Feeling bad on facebook: depression disclosures by college students on a social networking site," *Depression and Anxiety*, vol. 28, no. 6, pp. 447–455, 2011.
- [24] J. A. Welsh, R. L. Nix, C. Blair, K. L. Bierman, and K. E. Nelson, "The development of cognitive skills and gains in academic school readiness for children from low-income families," *Journal of Educational Psychology*, vol. 102, no. 1, 2010.
- [25] F. Lievens and P. R. Sackett, "The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance," *Journal of Applied Psychology*, vol. 97, no. 2, 2012.
- [26] K. Murayama, R. Pekrun, S. Lichtenfeld, and R. Vom Hofe, "Predicting long-term growth in students' mathematics achievement: the unique contributions of motivation and cognitive strategies," *Child Development*, vol. 84, no. 4, pp. 1475–1490, 2013.
- [27] M. Komaraju, A. Ramsey, and V. Rinella, "Cognitive and non-cognitive predictors of college readiness and performance: role of academic discipline," *Learning and Individual Differences*, vol. 24, pp. 103–109, 2013.
- [28] C. Valiente, K. Lemery-Chalfant, and J. Swanson, "Prediction of kindergartners' academic achievement from their effortful control and emotionality: evidence for direct and moderated relations," *Journal of Educational Psychology*, vol. 102, no. 3, 2010.
- [29] R. Kabra and R. Bichkar, "Performance prediction of engineering students using decision trees," *International Journal of computer applications*, vol. 36, no. 11, pp. 8–12, 2011.
- [30] I. D. Oladipo, J. B. Awotunde, M. AbdulRaheem et al., "An improved course recommendation system based on historical grade data using logistic regression," in *Proceedings of the International Conference on Applied Informatics*, pp. 207–221, Springer, Buenos Aires, Argentina, October 2021.
- [31] V. Ramesh, P. Parkavi, and K. Ramar, "Predicting student performance: a statistical and data mining approach," *International journal of computer applications*, vol. 63, no. 8, 2013.
- [32] D. Kabakchieva, "Student performance prediction by using data mining classification algorithms," *International journal of computer science and management research*, vol. 1, no. 4, pp. 686–690, 2012.
- [33] D. Kabakchieva, "Predicting student performance by using data mining methods for classification," *Cybernetics and Information Technologies*, vol. 13, no. 1, pp. 61–72, 2013.
- [34] B. K. Bhardwaj and S. Pal, "Data mining: a prediction for performance improvement using classification," 2012, <https://arxiv.org/abs/1201.3418>.
- [35] S. K. Yadav and S. Pal, "Data mining: a prediction for performance improvement of engineering students using classification," 2012, <https://arxiv.org/abs/1203.3832>.
- [36] O. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, "Application of k means clustering algorithm for prediction of students academic performance," 2010, <https://arxiv.org/abs/1002.2425>.
- [37] N. Thai-Nghe, L. Drumond, A. Krohn-Grimberghe, and L. Schmidt-Thieme, "Recommender system for predicting student performance," *Procedia Computer Science*, vol. 1, no. 2, pp. 2811–2819, 2010.
- [38] A. S. Elden, M. A. Moustafa, H. M. Harb, and A. H. Emara, "Adaboost ensemble with simple genetic algorithm for student prediction model," *AIRCC's International Journal of Computer Science and Information Technology*, vol. 5, no. 2, pp. 73–85, 2013.
- [39] M. Mayilvaganan and D. Kalpanadevi, "Comparison of classification techniques for predicting the performance of students academic environment," in *Proceedings of the International Conference on Communication and Network Technologies*, pp. 113–118, IEEE, Sivakasi, India, December 2014.
- [40] C. Watson, F. W. Li, and J. L. Godwin, "Predicting performance in an introductory programming course by logging and analyzing student programming behavior," in *Proceedings of the IEEE 13th international conference on advanced learning technologies*, pp. 319–323, IEEE, Beijing, China, July 2013.
- [41] Z. A. Pardos, S. M. Gowda, R. S. Baker, and N. T. Heffernan, "The sum is greater than the parts: ensembling models of student knowledge in educational software," *ACM SIGKDD explorations newsletter*, vol. 13, no. 2, pp. 37–44, 2012.
- [42] W. Jiang, Z. A. Pardos, and Q. Wei, "Goal-based course recommendation," in *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pp. 36–45, March 2019.
- [43] J. Zhang, B. Hao, B. Chen, C. Li, H. Chen, and J. Sun, "Hierarchical reinforcement learning for course recommendation in moocs," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 435–442, 2019.
- [44] V. A. Nguyen, H.-H. Nguyen, D.-L. Nguyen, and M.-D. Le, "A course recommendation model for students based on learning outcome," *Education and Information Technologies*, vol. 26, pp. 1–27, 2021.
- [45] S. Rao, K. Salomatin, G. Polatkan et al., "Learning to be relevant: evolution of a course recommendation system," in *Proceedings of the 28th ACM International Conference on*

- Information and Knowledge Management*, pp. 2625–2633, November 2019.
- [46] D. Yao and X. Deng, “A course teacher recommendation algorithm based on improved latent factor model and personalrank,” *IEEE Access*, vol. 9, Article ID 108614, 2021.
- [47] W. Xu and Y. Zhou, “Course video recommendation with multimodal information in online learning platforms: a deep learning framework,” *British Journal of Educational Technology*, vol. 51, no. 5, pp. 1734–1747, 2020.
- [48] M. S. Nassr and S. S. Abu-Naser, “Its for enhancing training methodology for students majoring in electricity,” *International Journal of Academic Pedagogical Research (IJAPR)*, vol. 3, 2019.
- [49] I. Ismail, E. Elihami, and M. Mustakim, “Students’ perceptions of the benefits of mobile polling technology in teaching and learning in college: implications of students’ participation and academic performance,” *Jurnal Pendidikan Progresif*, vol. 9, no. 1, pp. 89–104, 2019.
- [50] M. Buil-Fabrega, M. Martínez Casanovas, and N. Ruiz-Munzón, “Flipped classroom as an active learning methodology in sustainable development curricula,” *Sustainability*, vol. 11, no. 17, p. 4577, 2019.
- [51] I. U. Muradilloevich, O. K. Tanzilovch, A. A. Anvarovich, and S. I. Baxodirovna, “Improvement of teaching methodology by using modeling programs of engineering education in higher education of Uzbekistan,” *Journal of Critical Reviews*, vol. 7, no. 14, pp. 81–88, 2020.
- [52] P. Häfner, V. Häfner, and J. Ovtcharova, “Teaching methodology for virtual reality practical course in engineering education,” *Procedia Computer Science*, vol. 25, pp. 251–260, 2013.
- [53] D. González-Gómez, J. S. Jeong, and D. A. Rodríguez, “Performance and perception in the flipped learning model: an initial approach to evaluate the effectiveness of a new teaching methodology in a general science classroom,” *Journal of Science Education and Technology*, vol. 25, no. 3, pp. 450–459, 2016.
- [54] N. David, *Language Teaching Methodology*, Printice Hall, New York, London, Toronto, Sydney, Tokyo, Singapore, 1991.
- [55] S. Mahony and E. Pierazzo, *Teaching Skills or Teaching Methodology?*, OpenBook Publishers, Cambridge, UK, 2012.
- [56] A. Vujaklija, D. Hren, D. Sambunjak et al., “Can teaching research methodology influence students’ attitude toward science? cohort study and nonrandomized trial in a single medical school,” *Journal of Investigative Medicine*, vol. 58, no. 2, pp. 282–286, 2010.
- [57] S. Ahmad, K. Li, H. A. I. Eddine, and M. I. Khan, “A biologically inspired cognitive skills measurement approach,” *Biologically inspired cognitive architectures*, vol. 24, pp. 35–46, 2018.
- [58] F. Yang and F. W. Li, “Study on student performance estimation, student progress analysis, and student potential prediction based on data mining,” *Computers & Education*, vol. 123, pp. 97–108, 2018.
- [59] R. Trakunphutthirak and V. C. Lee, “Application of educational data mining approach for student academic performance prediction using progressive temporal data,” *Journal of Educational Computing Research*, vol. 59, Article ID 07356331211048777, 2021.
- [60] S. Helal, J. Li, L. Liu et al., “Predicting academic performance by considering student heterogeneity,” *Knowledge-Based Systems*, vol. 161, pp. 134–146, 2018.
- [61] Z. Xu, H. Yuan, and Q. Liu, “Student performance prediction based on blended learning,” *IEEE Transactions on Education*, vol. 64, no. 1, pp. 66–73, 2020.
- [62] S. Ahmad, M. S. Anwar, M. Ebrahim et al., “Deep network for the iterative estimations of students’ cognitive skills,” *IEEE Access*, vol. 8, Article ID 103100, 2020.
- [63] S. Ahmad, K. Li, A. Amin, and S. Khan, “A novel technique for the evaluation of posterior probabilities of student cognitive skills,” *IEEE Access*, vol. 6, Article ID 53153, 2018.
- [64] A. Acharya and D. Sinha, “Early prediction of students performance using machine learning techniques,” *International Journal of Computer Application*, vol. 107, no. 1, 2014.
- [65] A. A. Saa, “Educational data mining & students’ performance prediction,” *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, pp. 212–220, 2016.
- [66] C. Anuradha and T. Velmurugan, “A comparative analysis on the evaluation of classification algorithms in the prediction of students performance,” *Indian Journal of Science and Technology*, vol. 8, no. 15, pp. 1–12, 2015.
- [67] T. Mishra, D. Kumar, and S. Gupta, “Mining students’ data for prediction performance,” in *Proceedings of the 4th International Conference on Advanced Computing & Communication Technologies*, pp. 255–262, IEEE, Rohtak, India, February 2014.
- [68] S. Ahmad, K. Li, A. Amin, M. S. Anwar, and W. Khan, “A multilayer prediction approach for the student cognitive skills measurement,” *IEEE Access*, vol. 6, Article ID 57470, 2018.
- [69] L. Magnussen, D. Ishida, and J. Itano, “The impact of the use of inquiry-based learning as a teaching methodology on the development of critical thinking,” *Journal of Nursing Education*, vol. 39, 2000.
- [70] P. Tragazikis and M. Meimaris, “Engaging kids with the concept of sustainability using a commercial video game—a case study,” in *Transactions on Edutainment III*, pp. 1–12, Springer, Berlin, Germany, 2009.
- [71] D. H. Solomon, B. Lu, Z. Yu et al., “Benefits and sustainability of a learning collaborative for implementation of treat-to-target in rheumatoid arthritis: results of a cluster-randomized controlled phase ii clinical trial,” *Arthritis Care & Research*, vol. 70, no. 10, pp. 1551–1556, 2018.
- [72] P. E. Waggoner and J. H. Ausubel, “A framework for sustainability science: a renovated ipat identity,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7860–7865, 2002.
- [73] M. A. Van Waas, “Determinants of dissatisfaction with dentures: a multiple regression analysis,” *The Journal of Prosthetic Dentistry*, vol. 64, no. 5, pp. 569–572, 1990.
- [74] Y. Lee, M. L. Wehmeyer, S. B. Palmer, K. Williams-Diehm, D. K. Davies, and S. E. Stock, “Examining individual and instruction-related predictors of the self-determination of students with disabilities: multiple regression analyses,” *Remedial and Special Education*, vol. 33, no. 3, pp. 150–161, 2012.
- [75] R. L. Prentice, “Correlated binary regression with covariates specific to each binary observation,” *Biometrics*, vol. 44, pp. 1033–1048, 1988.
- [76] L. C. Soodak and D. M. Podell, “Teacher efficacy and student problem as factors in special education referral,” *The Journal of Special Education*, vol. 27, no. 1, pp. 66–81, 1993.
- [77] C. G. Thompson, R. S. Kim, A. M. Aloe, and B. J. Becker, “Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results,” *Basic and Applied Social Psychology*, vol. 39, no. 2, pp. 81–90, 2017.

- [78] M. Li, Y. Chen, C. Lal, M. Conti, M. Alazab, and D. Hu, "Eunomia: anonymous and secure vehicular digital forensics based on blockchain," *IEEE Transactions on Dependable and Secure Computing*, 2021.
- [79] M. Li, Y. Chen, S. Zheng, D. Hu, C. Lal, and M. Conti, "Privacy-preserving navigation supporting similar queries in vehicular networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, 2020.
- [80] S. Khan, Z. Zhang, L. Zhu, M. Li, Q. G. Khan Safi, and X. Chen, "Accountable and transparent tls certificate management: an alternate public-key infrastructure with verifiable trusted parties," *Security and Communication Networks*, vol. 2018, Article ID 8527010, 16 pages, 2018.
- [81] S. Khan, L. Zhu, X. Yu et al., "Accountable credential management system for vehicular communication," *Vehicular Communications*, vol. 25, Article ID 100279, 2020.
- [82] Z. Zhang, M. Li, L. Zhu, and X. Li, "Smartdetect: a smart detection scheme for malicious web shell codes via ensemble learning," in *Proceedings of the International Conference on Smart Computing and Communication*, pp. 196–205, Springer, Tokyo, Japan, December 2018.
- [83] C. F. Rodríguez-Hernández, M. Musso, E. Kyndt, and E. Cascallar, "Artificial neural networks in academic performance prediction: systematic implementation and predictor evaluation," *Computers & Education: Artificial Intelligence*, vol. 2, Article ID 100018, 2021.
- [84] I. E. Livieris, K. Drakopoulou, and P. Pintelas, "Predicting students' performance using artificial neural networks," in *Proceedings of the 8th PanHellenic Conference with International Participation Information and Communication Technologies in Education*, pp. 321–328, Volos, Greece, September 2012.
- [85] Z. Hu, Y. Xing, C. Lv, P. Hang, and J. Liu, "Deep convolutional neural network-based Bernoulli heatmap for head pose estimation," *Neurocomputing*, vol. 436, pp. 198–209, 2021.
- [86] B. Naik and S. Ragothaman, "Using neural networks to predict mba student success," *College Student Journal*, vol. 38, no. 1, pp. 143–150, 2004.
- [87] S. S. Abu-Naser, I. S. Zaqout, M. Abu Ghosh, R. R. Atallah, and E. Alajrami, "Predicting student performance using artificial neural network," *In the faculty of engineering and information technology*, vol. 8, 2015.
- [88] R. L. U. Cazarez and C. L. Martin, "Neural networks for predicting student performance in online education," *IEEE Latin America Transactions*, vol. 16, no. 7, pp. 2053–2060, 2018.
- [89] T. Gedeon and H. Turner, "Explaining student grades predicted by a neural network," vol. 1, pp. 609–612, in *Proceedings of the 1993 International Conference on Neural Networks (IJCNN-93)*, vol. 1, IEEE, Nagoya, Japan, October 1993.
- [90] L. H. Son and H. Fujita, "Neural-fuzzy with representative sets for prediction of student performance," *Applied Intelligence*, vol. 49, no. 1, pp. 172–187, 2019.
- [91] Y. Kara, M. A. Boyacioglu, and Ö. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the istanbul stock exchange," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5311–5319, 2011.
- [92] Y. Li and W. Ma, "Applications of artificial neural networks in financial economics: a survey," in *International symposium on computational intelligence and design*, vol. 1, pp. 211–214, IEEE, 2010.
- [93] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," 2012, <https://arxiv.org/abs/1201.3417>.
- [94] S. Huang and N. Fang, "Predicting student academic performance in an engineering dynamics course: a comparison of four types of predictive mathematical models," *Computers & Education*, vol. 61, pp. 133–145, 2013.
- [95] C. Romero, P. G. Espejo, A. Zafra, J. R. Romero, and S. Ventura, "Web usage mining for predicting final marks of students that use moodle courses," *Computer Applications in Engineering Education*, vol. 21, no. 1, pp. 135–146, 2013.
- [96] Z. Hu, Y. Zhang, Y. Xing, Y. Zhao, D. Cao, and C. Lv, *Towards Human-Centered Automated Driving: A Novel Spatial-Temporal Vision Transformer-Enabled Head Tracker*, 2022.
- [97] C. Romero, M.-I. López, J.-M. Luna, and S. Ventura, "Predicting students' final performance from participation in on-line discussion forums," *Computers & Education*, vol. 68, pp. 458–472, 2013.
- [98] S. Kotsiantis, K. Patriaracheas, and M. Xenos, "A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education," *Knowledge-Based Systems*, vol. 23, no. 6, pp. 529–535, 2010.
- [99] O. A. Echeagaray-Calderon and D. Barrios-Aranibar, "Optimal selection of factors using genetic algorithms and neural networks for the prediction of students' academic performance," in *Proceedings of the Latin America Congress on Computational Intelligence (LA-CCI)*, pp. 1–6, IEEE, Curitiba, Brazil, October 2015.
- [100] A. Siri, "Predicting students' dropout at university using artificial neural networks," *Italian Journal of Sociology of Education*, vol. 7, no. 2, 2015.
- [101] S. H. Teshnizi and S. M. T. Ayatollahi, "A comparison of logistic regression model and artificial neural networks in predicting of student's academic failure," *Acta Informatica Medica*, vol. 23, no. 5, p. 296, 2015.
- [102] M. Saarela and T. Kärkkäinen, "Analysing student performance using sparse data of core bachelor courses," *Journal of educational data mining*, vol. 7, no. 1, 2015.
- [103] K. Shaleena and S. Paul, "Data mining techniques for predicting student performance," in *Proceedings of the IEEE international conference on engineering and technology (ICETECH)*, pp. 1–3, IEEE, Coimabto, India, March 2015.
- [104] R. S. Agrawal and M. H. Pandya, "Survey of papers for data mining with neural networks to predict the student's academic achievements," *International Journal of Computer Science Trends and Technology (IJCSST)*, vol. 3, p. I5, 2015.
- [105] D. K. Kolo and S. A. Adepoju, "A decision tree approach for predicting students academic performance," *International Journal of Education and Management Engineering*, vol. 5, 2015.
- [106] R. Suchithra, V. Vaidhehi, and N. E. Iyer, "Survey of learning analytics based on purpose and techniques for improving student performance," *International Journal of Computer Application*, vol. 111, no. 1, 2015.
- [107] V. Bansal, H. Buckchash, and B. Raman, "Computational intelligence enabled student performance estimation in the age of covid-19," *SN computer science*, vol. 3, no. 1, pp. 1–11, 2022.
- [108] S. Gupta and N. Mishra, "Artificial intelligence and deep learning-based information retrieval framework for assessing student performance," *International Journal of Information Retrieval Research*, vol. 12, no. 1, pp. 1–27, 2022.

- [109] M. von Davier, L. Tyack, and L. Khorramdel, "Automated scoring of graphical open-ended responses using artificial neural networks," 2022, <https://arxiv.org/abs/2201.01783>.
- [110] R. M. Ali S and S. Perumal, "Multi-class lda classifier and cnn feature extraction for student performance analysis during covid-19 pandemic," *International Journal of Nonlinear Analysis and Applications*, vol. 13, no. 1, pp. 1329–1339, 2022.
- [111] A. P. Fard and M. H. Mahoor, "Facial landmark points detection using knowledge distillation-based neural networks," *Computer Vision and Image Understanding*, vol. 215, Article ID 103316, 2022.
- [112] L. T. Yogarathinam, K. Velswamy, A. Gangasalam et al., "Performance evaluation of whey flux in dead-end and cross-flow modes via convolutional neural networks," *Journal of Environmental Management*, vol. 301, Article ID 113872, 2022.
- [113] Z. Hu, C. Lv, P. Hang, C. Huang, and Y. Xing, "Data-driven estimation of driver attention using calibration-free eye gaze and scene features," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 2, pp. 1800–1808, 2021.
- [114] Y. Jedidi, A. Ibriz, M. Benslimane, M. Tmimi, and M. Rahhali, "Predicting student's performance based on cloud computing," in *Proceedings of the 6th International Conference on Wireless Technologies, Embedded, and Intelligent Systems WITS*, pp. 113–123, Springer, July 2022.
- [115] A. Roy, M. Rahman, M. N. Islam, N. I. Saimon, M. Alfaz, and A.-A.-S. Jaber, "A deep learning approach to predict academic result and recommend study plan for improving student's academic performance," in *Ubiquitous Intelligent Systems*, pp. 253–266, Springer, Berlin, Germany, 2022.
- [116] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, pp. 1–10, June 2008.
- [117] S. Keele, *Guidelines for Performing Systematic Literature Reviews in Software Engineering*, 2007.
- [118] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of Systems and Software*, vol. 80, no. 4, pp. 571–583, 2007.
- [119] L. Kennelly and M. Monrad, *Easing the Transition to High School: Research and Best Practices Designed to Support High School Learning*, National High School Center, 2007.
- [120] L. Kennelly and M. Monrad, *Approaches to Dropout Prevention: Heeding Early Warning Signs with Appropriate Interventions*, American Institutes for Research, Virginia, DC, USA, 2007.
- [121] C. M. Rodriguez, L. R. Baker, D. F. Pu, and M. C. Tucker, "Predicting parent-child aggression risk in mothers and fathers: role of emotion regulation and frustration tolerance," *Journal of Child and Family Studies*, vol. 26, no. 9, pp. 2529–2538, 2017.
- [122] B. Griffin and W. Hu, "Parental career expectations: effect on medical students' career attitudes over time," *Medical Education*, vol. 53, no. 6, pp. 584–592, 2019.
- [123] B. A. Trammell and C. LaForge, "Common challenges for instructors in large online courses: strategies to mitigate student and instructor frustration," *Journal of Educators Online*, vol. 14, no. 1, 2017.
- [124] M. L. George, "Effective teaching and examination strategies for undergraduate learning during covid-19 school restrictions," *Journal of Educational Technology Systems*, vol. 49, no. 1, pp. 23–48, 2020.
- [125] D. Carless and N. Winstone, "Teacher feedback literacy and its interplay with student feedback literacy," *Teaching in Higher Education*, vol. 25, pp. 1–14, 2020.
- [126] W. Wan Chik, Y. Salamonson, B. Everett et al., "Gender difference in academic performance of nursing students in a malaysian university college," *International Nursing Review*, vol. 59, no. 3, pp. 387–393, 2012.

Research Article

Intelligent Monitoring System Based on Noise-Assisted Multivariate Empirical Mode Decomposition Feature Extraction and Neural Networks

Le Fa Zhao ¹, Shahin Siahpour ², Mohammad Reza Haeri Yazdi ³, Moosa Ayati ³,
and Tian Yu Zhao ⁴

¹School of General Education, Shenyang Sport University, Shenyang 110115, China

²Department of Mechanical Engineering, University of Cincinnati, Cincinnati 45221, USA

³School of Mechanical Engineering, University of Tehran, Tehran, Iran

⁴Key Laboratory of Structural Dynamics of Liaoning Province, College of Sciences, Northeastern University, Shenyang 110819, China

Correspondence should be addressed to Le Fa Zhao; larry2012@syty.edu.cn and Tian Yu Zhao; zhaotianyu@mail.neu.edu.cn

Received 4 March 2022; Accepted 31 March 2022; Published 25 April 2022

Academic Editor: Jie Liu

Copyright © 2022 Le Fa Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Because of the nonlinearity and nonstationarity in the vibration signals of some rotating machinery, the analysis of these signals using conventional time- or frequency-domain methods has some drawbacks, and the results can be misleading. In this paper, a couple of features derived from multivariate empirical mode decomposition (MEMD) are introduced, which overcomes the shortcomings of the traditional features. A wind turbine gearbox and its bearings are investigated as rotating machinery. In this method, two types of feature structures are extracted from the decomposed signals resulting from the MEMD algorithm, called intrinsic mode function (IMF). The first type of feature vector element is the energy moment of effective IMFs. The other type of vector elements is amplitudes of a signal spectrum at the characteristic frequencies. A correlation factor is used to detect effective IMFs and eliminate the redundant IMFs. Since the basic MEMD algorithm is sensitive to noise, a noise-assisted extension of MEMD, NA-MEMD, is exploited to reduce the effect of noise on the output results. The capability of the proposed feature vector in health condition monitoring of the system is evaluated and compared with traditional features by using a discrimination factor. The proposed feature vector is utilized in the input layer of the classical three-layer backpropagation neural network. The results confirm that these features are appropriate for intelligent fault detection of complex rotating machinery and can diagnose the occurrence of early faults.

1. Introduction

With the advent of new era of Industry 4.0, the human and machine interaction has dramatically changed [1]. The improvement and advancement in intelligent systems have paved the way for the better use of smart devices. This shifts traditional human-machine interactions (HMI) toward intelligent human-machine interactions. The application of intelligent HMI ranges from medical scenarios to industrial applications [2–5] (e.g., robotics, energy, maintenance, and semiconductor

manufacturing). Among the key drivers of the transition from traditional to intelligent HMI, progress in machine learning and intelligent algorithms constitutes the main portion of importance [6–9].

Monitoring the condition of rotating machinery plays an important role in the engineering industries [10, 11]. To detect early faults and fully inspect the health condition of rotating systems, a condition monitoring structure is required to operate as soon as possible [12, 13]. The main objective of exploiting condition monitoring systems is to improve accuracy by lowering costs. The extraction of fault

characteristics from these types of systems is a key step in the process of fault detection and condition monitoring [14].

Signals from complex rotating machinery are usually nonstationary and nonlinear, and extracting features that lead to a desirable outcome has become a challenging process. Features are the parameters that are derived from signals to indicate the characteristics of systems. So far, various features that can be extracted from vibration signals have been investigated [15–17]. Signal processing to extract fault features is divided into three main domains: time domain, frequency domain, and time-frequency domain. Some conventional time-domain methods are skewness and kurtosis [18] or root mean square (RMS) and peak value of a signal [19]. Frequency analysis mostly contains Fourier spectra of a time series signal, cepstrum analysis, or envelope analysis [20, 21]. These features are in the time or frequency domain and are mostly extracted from raw vibration signals. In the presence of nonlinearity and nonstationarity in the signal, traditional features cannot have an accurate distinction between system conditions [22]. Because of these problems, time-frequency analysis of complex signals is introduced as an application of feature extraction. Time-frequency methods, such as the short-time Fourier transform [23], wavelet transform [24], empirical mode decomposition (EMD) [25], or Wigner–Ville [26], analyze signals in both time and frequency domains. Therefore, features can contain more comprehensive information of signals.

With the advent of a new time-frequency method, named Hilbert–Huang transform (HHT) [27], many studies have been conducted using this method in the field of signal processing [28–30]. HHT is a powerful algorithm useful for nonlinear and nonstationary signals, performing an adaptive decomposition operation called empirical mode decomposition (EMD). The decomposed signals, named intrinsic mode functions (IMFs), are almost monocomponents which satisfy Hilbert transform terms. Each IMF covers a small range of frequency scales. This characteristic of IMFs makes them a suitable tool for the analysis of complex systems. EMD algorithm is sensitive to noise. When signals are noisy, the mode-mixing phenomenon can occur in IMFs [31]. In this situation, either a single IMF carries a signal of a widely disparate scale, or a single mode (or scaling) exists in more than one IMF. To overcome this phenomenon, Ensemble EMD (EEMD) is proposed [32].

When the system contains many components and has comprehensive information from all over the system, multiple sensors are located on different parts of the system. In this condition, the signals obtained from the sensors are a kind of multivariate signals. If the EMD algorithm is used on each signal individually, joint information will be wasted [33]. Furthermore, the same group of IMFs may have different characteristic information [34]. To overcome these problems, Riling et al. [35] proposed bivariate EMD. In this method, by mapping the bivariate signal in different directions, the local mean of the signal is calculated. To continue this idea, in 2010, Rehman and Mandic [36] proposed an empirical mode decomposition algorithm for trivariate signals. After that, they proposed an extension to

their method and introduced multivariate EMD (MEMD) to deal with multidimensional signals [37]. This method allows us to analyze multidimensional signals simultaneously and covers the problem of using the EMD method for these kinds of signals. Zhao et al. [38] employ multivariate EMD method to extract some health condition information of the studied system. In their study, they used full spectrum based condition monitoring for rotating machinery. Lv et al. [33] used multivariate EMD as an application to investigate the health conditions of the patients.

Each IMF order resulting from the MEMD algorithm has the same frequency characteristic. This capability makes the MEMD algorithm a suitable method for feature extraction to diagnose faults in rotating systems. Some of the IMFs are spurious and need to be eliminated from the calculation to speed up the process of feature extraction and make the feature vector smaller without losing accuracy. Some IMFs are high-frequency ones, which can be regarded as noisy IMFs. In contrast, some IMFs contain low-frequency characteristics that exist due to the stopping criteria of the EMD algorithm and do not have physical meaning. Effective IMFs can be detected by user experience, but to make the process faster, a criterion or factor must be used. Ricci et al. [39] introduced a merit index that automatically selects the effective IMFs and eliminates the spurious ones. This index is based on the symmetrical and periodic IMF specifications. In [38], a sensitivity factor which is based on mutual information is proposed. In [33], a correlation factor is introduced to detect the most effective IMFs and, as is obvious from the name of the factor, it is based on the correlation between the signal and each IMF.

The features derived from the signals can be implemented as input for an artificial neural network (ANN) system [40] or can be used for a support vector machine (SVM) [41] to analyze the conditions of the system intelligently and automatically. Yang et al. [42] extract bearing health characteristics using the energy of decomposed IMFs. They compare the output results from a simple ANN while the features are derived from wavelet analysis. Bin et al. [43] used a combined method of wavelet packet decomposition (WPD) and EMD to extract fault features from a bearing mechanism as rotating machinery. In their study, the energy moment from the IMFs is used as the feature vector. WPD is used to denoise and preprocess a signal.

To address the aforementioned issues and challenges, an intelligent feature extraction is proposed. The following are the main novelties and contributions of this study:

- (i) The NA-MEMD algorithm is used as a feature extraction method.
- (ii) Correlation analysis is used to detect effective IMFs.
- (iii) In addition to the energy moment of effective IMFs, an amplitude factor in the frequency domain is introduced as a complementary element for the feature vector.
- (iv) To show the capability of the proposed features in the diagnosis of system conditions, a discrimination criterion is exploited to make the comparison

tangible. Features are then used for a back-propagation (BP) neural network input layer.

- (v) The proposed algorithm can be used for analyzing the features of the data from the athletes and the fault analysis of the key mechanical components in the sport field. This paper focuses on the analysis of bearings used in the key components in the sport field.

This paper is organized as follows. In Section 2, the proposed signal processing and feature extraction procedure are explained. Section 3 is dedicated to the structure and design configuration of the neural network. In Section 4, the rotation system is introduced. In Section 5, the implementation of the proposed method on the studied system is investigated, and the results are discussed. The conclusion is presented in Section 6.

2. Feature Extraction Using Multivariate EMD

2.1. Fundamentals of Multivariate EMD. In standard EMD [27], the local mean can be calculated by interpolating the upper and lower envelope of a univariate signal. However, when dealing with multivariate signals, it is rather confusing to determine IMFs, because the value of local minima and maxima cannot be directly defined. Rehman and Mandic [37] introduce a multivariate EMD algorithm to overcome these issues. In this method, multivariate (n -variate) signals are considered as n -dimensional time series. Some appropriate direction vectors are chosen, and multivariate signals are projected on the selected direction vectors. All envelopes of these projected signals are calculated, and by averaging the envelopes, the local mean of the multivariate signal is determined. Therefore, the sifting process [31] (which is used in standard EMD) can be implemented to calculate IMF groups.

The process of local mean calculation can be considered as an approximation of the integral of all envelopes along with the multiple directions in the n -dimensional space. The accuracy of this calculation depends on the uniformity of the chosen direction vectors. To generate a set of uniformly distributed points, quasi-Monte Carlo-based low-discrepancy sequences can be utilized. The Halton sequence family is exploited as a convenient way to generate a low-discrepancy sequence.

Let x_1, \dots, x_n be the first n prime numbers, and the i th sample of a one-dimensional Halton sequence, denoted by r_i^x , is given by

$$r_i^x = \frac{a_0}{x} + \frac{a_1}{x^2} + \dots + \frac{a_s}{x^{s+1}}, \quad (1)$$

where the base- x representation of i is given by

$$i = a_0 + a_1x + \dots + a_sx^s. \quad (2)$$

Starting from $i = 0$, the i th sample of Halton sequence then becomes

$$(r_i^{x_1}, r_i^{x_2}, \dots, r_i^{x_n}). \quad (3)$$

The Hammersley sequence is used when the total number of samples, n , is known a priori; in this case, the i th sample within the Hammersley sequence is calculated as

$$\left(\frac{i}{n}, r_i^{x_1}, r_i^{x_2}, \dots, r_i^{x_{n-1}} \right). \quad (4)$$

By using Halton and Hammersley sequences, a suitable set of direction vectors on the n -sphere is generated. Henceforth, projections of signals on this direction vector will be calculated. In the following paragraph, multivariate EMD will be explained briefly.

Let $X(t) = [x_1(t), x_2(t), \dots, x_n(t)]$ be an n -dimensional signal and $D^k = \{d_1^k, d_2^k, \dots, d_n^k\}$ correspond to the k th direction vector in a direction set D . The multivariate EMD algorithm is described as follows:

- (1) Choose a suitable set of direction vectors, D .
- (2) Calculate the k th projection, $p^k(t)$ of X along the k th direction, where $k = 1, 2, \dots, K$ and K is a total number of direction vectors.
- (3) Find the time instants, t_i^k , corresponding to the maxima of projected signals.
- (4) Interpolate $[t_i^k, X(t_i^k)]$ to determine multidimensional envelopes, $E^k(t)$.
- (5) Calculate the mean by

$$M(t) = \frac{1}{l} \sum_{k=0}^K E^k(t). \quad (5)$$

- (6) Calculate the residual component $R(t) = X(t) - M(t)$. If $D(t)$ satisfies the stopping criterion explained in the previous section, then consider $R(t)$ as an IMF and then repeat the algorithm until it meets the criterion.

2.2. Effect of Noise on IMFs. EMD method is sensitive to noise. In [44], an investigation is conducted on the sensitivity of MEMD to noise. It can be inferred from this study that the MEMD algorithm is sensitive to noise and mode-mixing problems that can happen in this method. An extension to MEMD is proposed to cover the problem. The extension is named noise-assisted multivariate empirical mode decomposition (NA-MEMD). NA-MEMD algorithm tries to eliminate noise interference in EEMD and reduce mode mixing in EMD and MEMD methods. The general algorithm in NA-MEMD is the same as in MEMD. The difference is that the input multivariate signal consists of input data and noise in separate channels. After the implementation of the MEMD algorithm on the new multivariate signal, the resulting noise-related IMFs will be discarded. This method is demonstrated briefly as follows:

- (1) Construct l -channel of uncorrelated Gaussian white noise time series which have the same length as that of the input ($l \geq 1$).
- (2) Add noise channels, created in the previous step, to the input signals; therefore, the new input signal is $(n+l)$ -channel.
- (3) Process the $(n+l)$ -channel multivariate signal using MEMD algorithm to obtain IMFs.

- (4) Discard l -channels corresponding to the noise from $(n+l)$ -variate IMFs and get n -channel IMFs corresponding to the original signal.

2.3. The Criterion for Choosing IMFs. To extract fault features from the signal, suitable IMFs must be selected. A suitable IMF is an IMF which has a meaningful frequency scale. The choice of IMF is usually based on experience and is done manually. This process is slow and time-consuming. To make this procedure faster and relatively automatic, an index or coefficient is needed to be introduced. One way to determine the suitability of an IMF is to calculate the correlation between the IMF and the original signal [45]. The IMF, which has a small correlation coefficient, is regarded as a redundant or noise component. With the help of the correlation coefficient, it is possible to accurately determine and eliminate the noise component and evaluate the effective IMFs to extract fault features from them.

In dealing with the MEMD algorithm, the resulting IMFs are a set of IMF groups, and some calculation must be done to identify the effective IMFs. Hence, a fault correlation factor (FCF) has been proposed [33] to conduct the analysis. Suppose that the input signal is n -variate signal and there exist n groups for m th IMFs corresponding to each signal. The multivariate signal can be organized as a matrix as follows:

$$S(t) = [S_1(t), S_2(t), \dots, S_n(t)]. \quad (6)$$

The k th IMF on n groups corresponds to each input signal and constitutes a matrix in the form of

$$C(t) = [c_1^k(t), c_2^k(t), \dots, c_n^k(t)]. \quad (7)$$

A simplified form of the correlation coefficient is as follows:

$$\lambda_{xy} = \frac{\sum_{n=1}^N x(t)c(t)}{\sqrt{\sum_{n=1}^N x^2(t) \sum_{n=1}^N c^2(t)}}, \quad (8)$$

where t is the time and N is the total number of sampling points. λ_i^k is defined as the FCF of i th IMF of $C(t)$ (7) and can be calculated by conducting correlation analysis on this IMF with each n -variate signal, respectively, and averaging all correlation factors. λ_i^k indicates the correlation between this IMF and the original signal. To make a comparison between each order of IMFs, the FCF of IMFs with the same order must be calculated. It can be achieved by averaging all vector correlations since each order of IMFs contains almost the same features.

$$\lambda^k = \sum_{i=1}^n \frac{\lambda_i^k}{n}. \quad (9)$$

When the value of λ^k is large, it means that the degree of correlation of the fault characteristic between the k th order IMF of the n IMF groups and the original signal is higher. Based on the criterion of Pearson Correlation Coefficients, when the value of the correlation coefficient is higher than

0.3, it can be assumed that the signals are relevant. Therefore with this approach, effective IMFs can be determined.

2.4. Feature Selection. The idea of extracting features for the diagnosis of rotating machinery faults is a critical task. Features must be selected wisely, because some features may be futile in extracting fault characteristics of a signal, although these parameters are useful for other vibration signals. To choose the most effective features, a scientific criterion, which relates the features to the system condition, can be used. To achieve this purpose, in this paper, a discrimination criterion, denoted as J , is applied [46]. This criterion is based on the ratio between inter- and intra-variance. Suppose N features are extracted for a vibration signal with K class of system conditions. If $r_{k,n}$ is the n th feature of the k th class, the intraclass and interclass variance matrix of the average dispersion coefficients are given as follows:

$$S_{\text{intra}} = \frac{1}{K \times N} \sum_{k=1}^K \sum_{n=1}^N (r_{k,n} - \underline{\mu}_k)(r_{k,n} - \underline{\mu}_k)^t, \quad (10)$$

$$S_{\text{inter}} = \frac{1}{K} \sum_{k=1}^K \sum_{n=1}^N (r_k - \underline{\mu}_c)(r_k - \underline{\mu}_c)^t,$$

while the mean of feature vectors of the k th class is defined by $\underline{\mu}_k = (1/N) \sum_{n=1}^N r_{k,n}$ and the total mean of feature vectors of all classes is $\underline{\mu}_c = (1/N) \sum_{k=1}^K \underline{\mu}_k$.

Finally, J is computed as follows:

$$J = \text{trace}(S_{\text{intra}}^{-1} S_{\text{inter}}). \quad (11)$$

According to the criterion, for the features with a high value of J , the effect of the corresponding feature on the diagnosis of a specific fault becomes greater.

2.5. Traditional Features. Traditional fault features are simple and can easily be implemented in signals [47]. In Table 1, some of these traditional features are represented in the frequency and time domain. When a fault occurs in the rotating machinery, the time-domain signal may change both its amplitude and distribution. Moreover, the frequency spectrum may encounter some deviation from the normal condition. Usually, with the help of these features, some faults can be determined in the system. Note. x_n is vibration signal with $n = 1, \dots, N$; N is the number of data points; s_k is the frequency spectrum of x_n ; K is number of spectral lines; and f_k is frequency value of k th spectral line.

2.6. Feature Extraction from Decomposed IMFs. In addition to the traditional features mentioned earlier, the MEMD algorithm is used to extract some other features to form a more reliable and almost more robust feature vector.

Standard EMD is designed to process univariate signals. When signals from multiple sensors (or conditions) are individually processed by the EMD algorithm, there might be two main drawbacks in the results. The first drawback is the loss of joint information. The main reason for collecting

TABLE 1: Traditional feature set parameters.

Time-domain features		Frequency-domain features	
Root mean square	$pt_1 = \sqrt{(1/N) \sum_{n=1}^N x_n^2}$	Frequency barycenter	$pf_1 = (\sum_{k=1}^K f_k s_k / \sum_{k=1}^K s_k)$
Peak	$pt_2 = \max(x_n)$	Root mean square frequency	$pf_2 = \sqrt{(\sum_{k=1}^K f_k^2 s_k / \sum_{k=1}^K s_k)}$
Square mean root	$pt_3 = ((1/N) \sum_{n=1}^N \sqrt{ x_n })^2$	Standard deviation frequency	$pf_3 = \sqrt{(\sum_{k=1}^K (f_k - pf_1)^2 s_k / \sum_{k=1}^K s_k)}$
Absolute mean	$pt_4 = (1/N) \sum_{n=1}^N (x_n)$	Frequency spectrum mean	$pf_4 = (1/N) \sum_{k=1}^K s_k$
Kurtosis	$pt_5 = (1/N) \sum_{n=1}^N x_n^4$	Frequency spectrum deviation	$pf_5 = (1/K - 1) \sum_{k=1}^K (s_k - pf_4)^2$
Crest factor	$pt_6 = (pt_1)/(pt_2)$	Frequency spectrum entropy	$pf_6 = -\sum_{k=1}^K (s_k/K pf_4) \log((s_k/K pf_4))$

information from multiple sensors (or conditions) is to have a more comprehensive understanding of the system. By implementation of EMD algorithm individually on each signal, the idea of multiple sensors would be vain. The second drawback is about the features of the same order of IMFs in each signal. IMFs in the same order corresponding to each signal that resulted from the EMD algorithm may have different features [34]. This makes it difficult to determine the effective IMFs.

MEMD algorithm overcomes these two problems. The IMFs, resulting from the MEMD algorithm, not only contain comprehensive information about the system, but also, in the same order of IMFs, almost consist of the same feature information. These two advantages of MEMD, in addition to the benefits of the EMD method, make this algorithm an ideal choice for extracting features contributing to multivariate signals.

As was mentioned before, each order of IMFs calculated by noise-assisted MEMD contains a small frequency scale. This characteristic paves the way for analysis and feature extraction in the frequency domain for each order of IMFs. When a fault occurs in a rotating component of a system, a natural frequency (or meshing frequency for contacting components, e.g., gearboxes) is excited, which results in a burst of energy at this frequency. To identify the fault, it is necessary to detect the frequency occurrence of these high-energy bursts. Since each IMF order is composed of a small range of frequencies, by performing frequency-domain analysis, the amplitude of the signal in characteristic frequencies can be determined. FCF is a suitable index to eliminate redundant IMFs or specifically redundant frequency bands. This amplitude can be regarded as a fault feature for implementation in smart analysis.

To clarify what was mentioned above, the procedure is implemented on the synthetic signal. The multivariate synthetic signal is given as follows:

$$\begin{aligned}
 x_1 &= \sin(2\pi f_1 t) + 0.5 \cos(2\pi f_2 t) + 0.9 \sin(2\pi f_3 t), \\
 x_2 &= 0.7 \sin(2\pi f_1 t) + \cos(2\pi f_2 t) + 0.4 \cos(2\pi f_2 t), \\
 x_3 &= 0.9 \sin(2\pi f_1 t) + 0.6 \cos(2\pi f_2 t) + \cos(2\pi f_2 t),
 \end{aligned} \quad (12)$$

where $f_1 = 20$ Hz, $f_2 = 50$ Hz, and $f_3 = 90$ Hz. The sampling point is $N = 1000$, and the sampling frequency is

$f_s = 1000$ Hz. White Gaussian noise is added to each signal. Noise signals are white Gaussian signals and the corresponding power is -10 dBW.

Since noise is added to the multivariate signal, to prevent the phenomenon of mode mixing, the NA-MEMD algorithm is implemented. Figure 1 shows the calculated IMFs by using NA-MEMD. From this figure, it is verified that each order of IMFs has the same frequency characteristics. IMF3 to IMF5 consist of the main frequencies of component signals. The remaining IMFs are redundant ones, either high-frequency IMFs which are regarded as noise or low-frequency IMFs which are due to the stopping criterion and do not have physical meaning.

FCF is used to determine effective IMFs and to detect which IMFs contain frequency features. In Table 2 the calculated results for FCF are shown. According to the criterion of Pearson Correlation Coefficients, since IMF3 to IMF5 have FCF higher than 0.3, they can be assumed to be relevant IMFs, which is acceptable for the manual estimation of these IMFs. Therefore, the process of selecting suitable IMFs converts to a relatively automatic procedure.

The amplitude of frequency spectrum of IMFs in characteristic frequencies is an ideal feature for fault detection of rotating machinery. In the dominant IMFs in the studied synthetic signal, there exist peaks in the propinquity of characteristic frequencies. The amplitude of these peaks is going to be used as a feature for the input of an artificial neural network, because the amplitude of these peaks changes when the system operates under different conditions. Therefore, this characteristic can make a distinction for different health conditions in the system.

To have accurate and reliable results from the neural network, features as the input of the neural network must contain the detailed information of the studied system. Vibration signals from rotating machinery are usually nonlinear and nonstationary. This specification of vibration signal, which changes the energy of the signal, is in some frequency bands. IMF components contain information corresponding to a frequency band; thus, the IMF energy can be used to characterize a signal. Instead of using energy [42] or the energy entropy of the signal [16], the energy moment [43] is used as part of the proposed characteristic vector. In this method, the time feature is used for the calculation of energy; thus, it can be a complementary feature extraction method in addition to the proposed

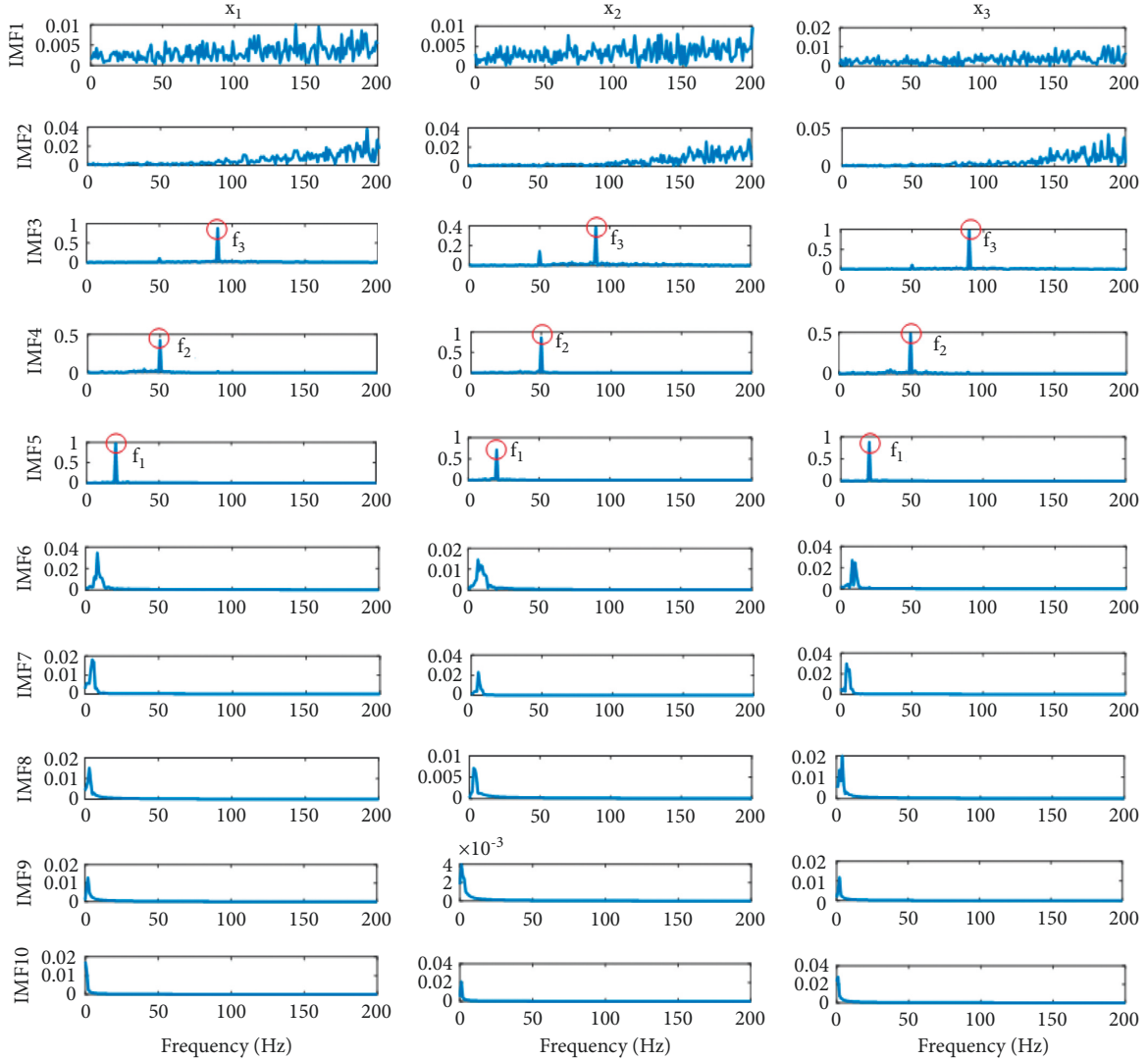


FIGURE 1: Decomposition results by using NA-MEMD on the synthetic multivariate signal.

TABLE 2: Fault correlation factor for synthetic multivariate signal.

IMF order	1	2	3	4	5	6	7	8	9	10
FCF	0.2131	0.1965	0.6491	0.4125	0.6600	0.0224	0.0208	0.0366	0.0376	0.0191

frequency-domain method. The energy moment can distinguish signal features more accurately compared to the classical energy method when the signal is nonlinear or nonstationary, which will be explained in the following paragraphs.

The energy moment for each IMF can be calculated as

$$E_i = \int t \cdot |c_i(t)|^2 dt, \quad (13)$$

and for continuous calculation and discrete analysis,

$$E_i = \sum_{k=1}^n (k\Delta t) |c_i(k\Delta t)|^2, \quad (14)$$

where n is the total number of sampling points, Δt is the period of samples, and k is the number of the sample points.

Energy moment can form a feature vector as follows:

$$T = [E_1, E_2, \dots, E_n]. \quad (15)$$

Because the energy moment has a high value, T can be adjusted using normalization. Assume $E = \sum_{i=1}^n E_i$; then,

$$T_n = \left[\frac{E_1}{E}, \frac{E_2}{E}, \dots, \frac{E_n}{E} \right], \quad (16)$$

where T_i is normalized energy moment for signal c_i . As is clear from (13) and (14), the moment energy contains both the signal energy and the signal distribution in the time domain (because of the term t in the equations), indicating the advantage of the moment energy over the calculation of the classical energy [43].

3. Neural Network Structure

A BP neural network is designed to intelligently diagnose faults in rotating machinery. To do so, a neural model of BP must be structured. A typical BP neural network structure is illustrated in Figure 2. This network has one hidden layer. In the field of rotating machinery fault detection, the input layer contains features extracted from the original signal, and the output layer is the system health conditions (i.e., being healthy or having a specific fault type).

The number of hidden layer cells cannot be defined accurately. If the hidden layer nodes are too high, the connection between nodes increases, and as a result, the number of connection weights increases, making the neural network training process more complex. If the hidden layer nodes are too small, the accuracy of the output results cannot be guaranteed. For a three-layer network (one hidden layer), there is an empirical and experimental relationship that relates the number of hidden layer nodes k to the number of input layer nodes n [43]. The relationship is given as follows:

$$k = 2n + i, 0 \leq i \leq 8. \quad (17)$$

Note that even in this relationship, k is not definite and can be changed.

In Figure 3, an overview of smart fault detection of rotating machinery is illustrated schematically.

4. System Description

To explain the proposed method, this paper investigates the transmission system in the wind turbine system (gearbox and bearing), as a rotating machinery. The vibration data from the experiment were provided by the National Renewable Energy Laboratory (NREL). The system is depicted in Figure 4. As is indicated in the figure, the main sections rotate at the three speed stages, i.e., the low-speed stage (LSS), the intermediate-speed stage (ISS), and the high-speed stage (HSS). The test drive is designed for the wind turbine with rated power of 750 kW. The overall ratio for the gearbox system is 1:81.491. In Table 3, more details on the description of the gearbox are shown [48].

To obtain vibration data from the gearbox system, accelerometers are mounted on the top of the gearbox. Data are collected at a rate of 40 kHz per channel using a National Instruments PXI-4472B high-speed data acquisition system (DAQ). Eight sensors are located in different places of the system to obtain comprehensive information from the gearbox system.

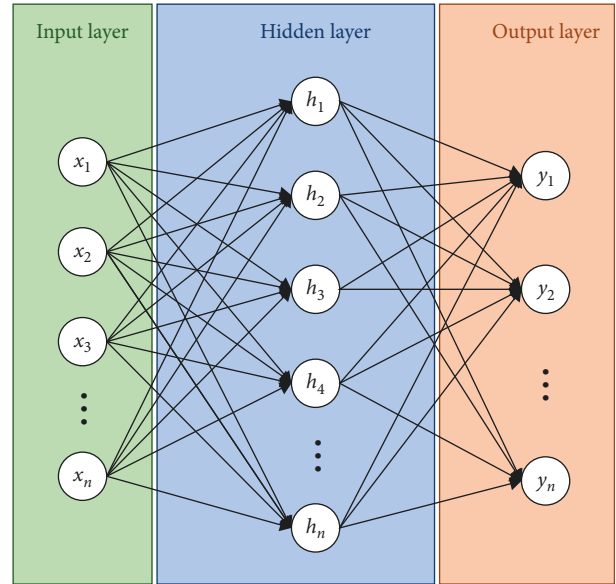


FIGURE 2: A typical BP neural network.

As was mentioned in the previous section, the feature vectors contain some components which are related to the amplitude of the frequency spectrum with the characteristic frequency. Characteristic frequency encompasses not only the rotating frequencies of the components but also the meshing frequencies of linked components. The studied system in faulty condition corresponds to three major fault types. The formulation for the calculation of the main characteristic frequencies is briefly illustrated in Table 4.

For the gearbox of fixed axis, $f_1, f_2, N_1,$ and N_2 are the frequency of the pinion, the frequency of the gear, the number of teeth in the pinion, and the number of teeth in the gear, respectively. For the planetary stage, $f_s, N_s, N_R,$ and N_p are the sun frequency, the number of suns, the ring gear, and the teeth of the planet, respectively. For the bearing, $f_r, n, \phi, d,$ and D are the shaft speed, the number of rolling elements, the angle of the load from the radial plane, the rolling element diameter, and the bearing average diameter, respectively. In Figure 5, the main dominant characteristic frequencies are shown schematically. These frequencies are high-speed shaft (HSS) frequency, planetary gear mesh frequency (PLTGM), high-speed shaft bearing B (Figure 4), high-speed shaft gear mesh (HSGM), and its second and third harmonics.

5. Method Implementation on System and Discussion

5.1. Feature Extraction for the System. As was mentioned, the input layer in the neural network is a vector constructed from fault features. Some elements are composed of normalized energy moments. First, a windowing process is implemented on the input signal to construct as many signals as possible for the input of the NA-MEMD algorithm as the input of the neural network. The signals provided by NREL are made up

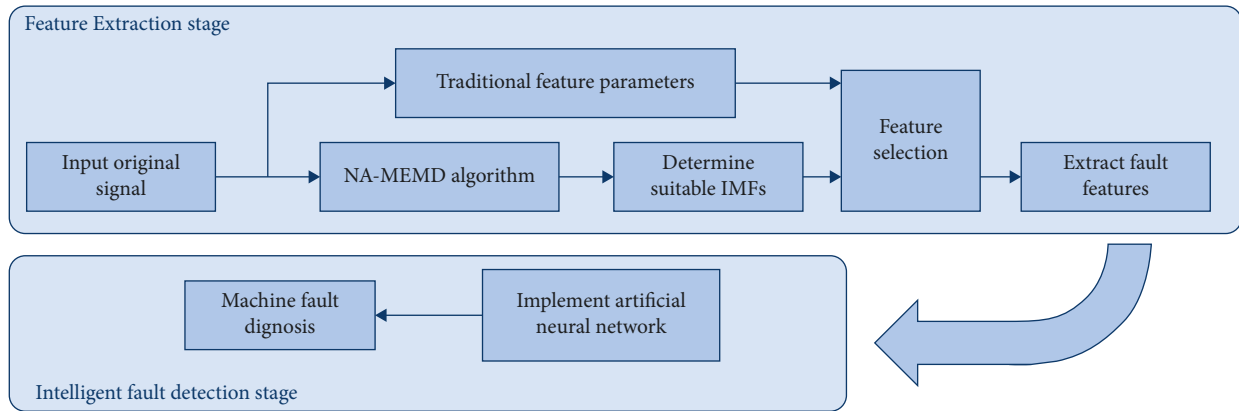


FIGURE 3: Intelligent fault detection flowchart.

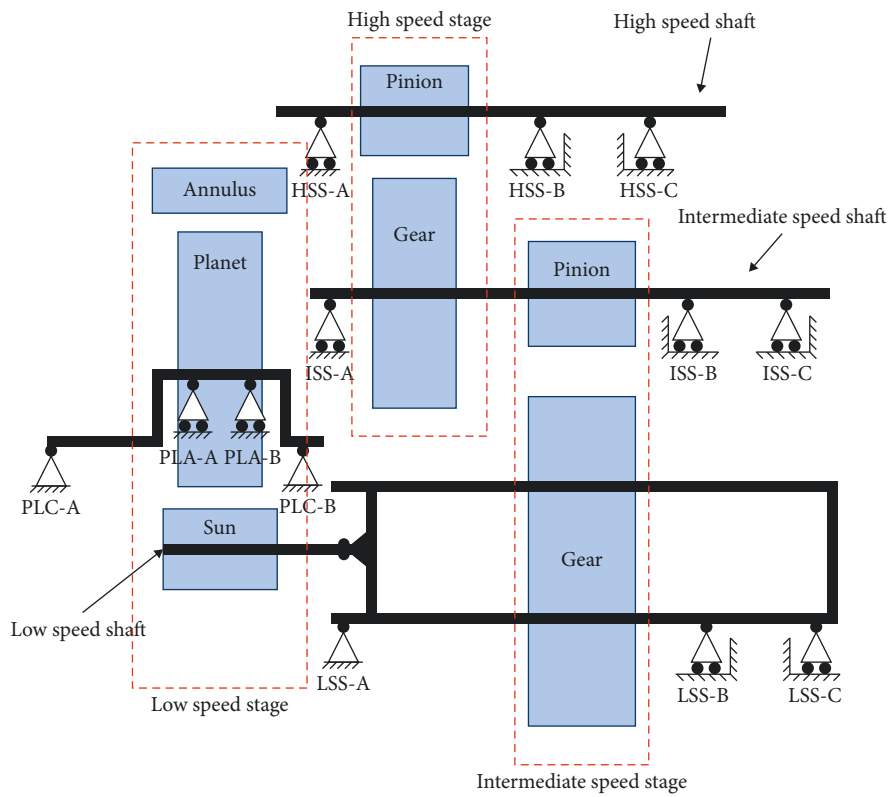


FIGURE 4: Wind turbine planetary gearbox system (courtesy of NREL).

TABLE 3: Dimensions and mechanical details of the gear element [48].

Gear Elements	No. of teeth	Mate teeth	Root diameter (mm)	Helix angle	Face width (mm)	Ratio
Ring gear	99	39	1047	7.5 L	230	
Planet gear	39	99	372	7.5 L	227.5	
Sun gear	21	39	186	7.5 R	220	5.71
Intermediate gear	82	23	678	14 R	170	
Intermediate pinion	23	82	174	14 L	186	3.57
High-speed gear	88	22	440	14 L	110	
High-speed pinion	22	88	100	14 R	120	4.0
					Overall:	81.49

TABLE 4: Characteristic frequencies formulations.

Component	Characteristic frequency	Formulation
Fixed-axis gearbox	Meshing frequency	$f_m = f_1 N_1 = f_2 N_2$
Planetary stage	Planet frequency [49]	$f_p = ((N_p - N_R)N_s / (N_R + N_s)N_p) f_s$
	Carrier frequency [49]	$f_c = (N_s / N_R + N_s) f_s$
	Meshing frequency [49]	$f_{m-p} = (f_s - f_c)N_s = (N_R N_s / N_R + N_s) f_s$
Bearing	Ball pass frequency, outer race [50]	$BPFO = (n f_r / 2) \{1 - (d/D) \cos \phi\}$
	Ball pass frequency, inner race [50]	$BPFI = (n f_r / 2) \{1 + (d/D) \cos \phi\}$
	Fundamental train frequency (cage speed) [50]	$FTF = (f_r / 2) \{1 - d/D \cos \phi\}$
	Ball (roller) spin frequency [50]	$BSF (RSF) = (D/2d) \{1 - ((d/D) \cos \phi)\}^2$

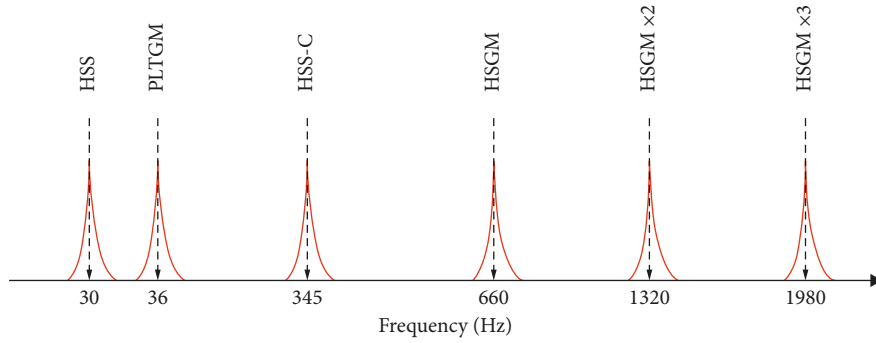


FIGURE 5: The characteristic frequency of the gearbox.

of 10 signals of 60 s duration. Each signal contains 2400000 data samples. A window is a section of each signal with 240000 data samples without overlapping which divides the corresponding signal into 10 subsignals. Thus, for each condition of the system, 100 features can be constructed. Windowing increases the feature vectors, increasing the accuracy of neural network operation. Subsections are now considered as input to the MEMD algorithm to obtain IMFs. To avoid the mode-mixing phenomenon, NA-MEMD is used instead of the MEMD algorithm. 3 white Gaussian noises with a variance of 0.1 are added as 3 new channels to the multivariate input signal. In Figure 6, the resulting IMFs for one channel of the multivariate faulty signal are shown. 20 IMFs are extracted from the NA-MEMD while some of them are spurious and must be omitted from the consideration. In Table 5, FCF values calculated for the IMFs are shown. IMFs of orders three to eight have an FCF higher than 0.3; thus, these IMF groups are considered as effective IMFs for the calculation of energy moment.

The feature selection algorithm is applied to the proposed features. For the system studied, two classes of system conditions are considered ($K = 2$) and 30 characteristics are extracted ($N = 30$). The resultant discrimination criterion is shown in Table 6. According to the table, the values of J for most of the MEMD characteristics are greater than the traditional characteristics except for the value of pt_1 (that is, the root mean square). This shows that the proposed features can be suitable for detecting faults in the wind turbine

gearbox studied. Therefore, the feature vector can be constructed as follows:

$$F = [E'_3, E'_4, E'_5, E'_6, E'_7, E'_8, AF_6, AF_8, AF_{11}], \quad (18)$$

where E'_i and AF_i are normalized energy moment and amplitude factor (AF) for the i th IMF order, respectively. It should be noted that these features are selected based on the studied dataset; however, the feature selection practice for all similar datasets is the same. It means the features with highest FCF value should be selected for the input of any machine learning method.

In Table 7 a feature vector as a sample is depicted. It can be seen from the table that the input vector is composed of nine features. Although the output layer of the neural network contains two conditions (i.e., healthy and faulty), it is worth noting that the faulty condition encompasses three different faults. Since the data provided consist of two conditions, inevitably two output conditions are chosen for the neural network. The trend of fault feature vectors is constructed to detect faults individually. However, in this paper, because of the limitation of data, faults are detected simultaneously in one condition label.

5.2. The Design Neural Network for the System. The main step in designing a neural network is to train the network based on the training samples. As mentioned earlier, the feature vectors in the input layer contain nine components. The number of

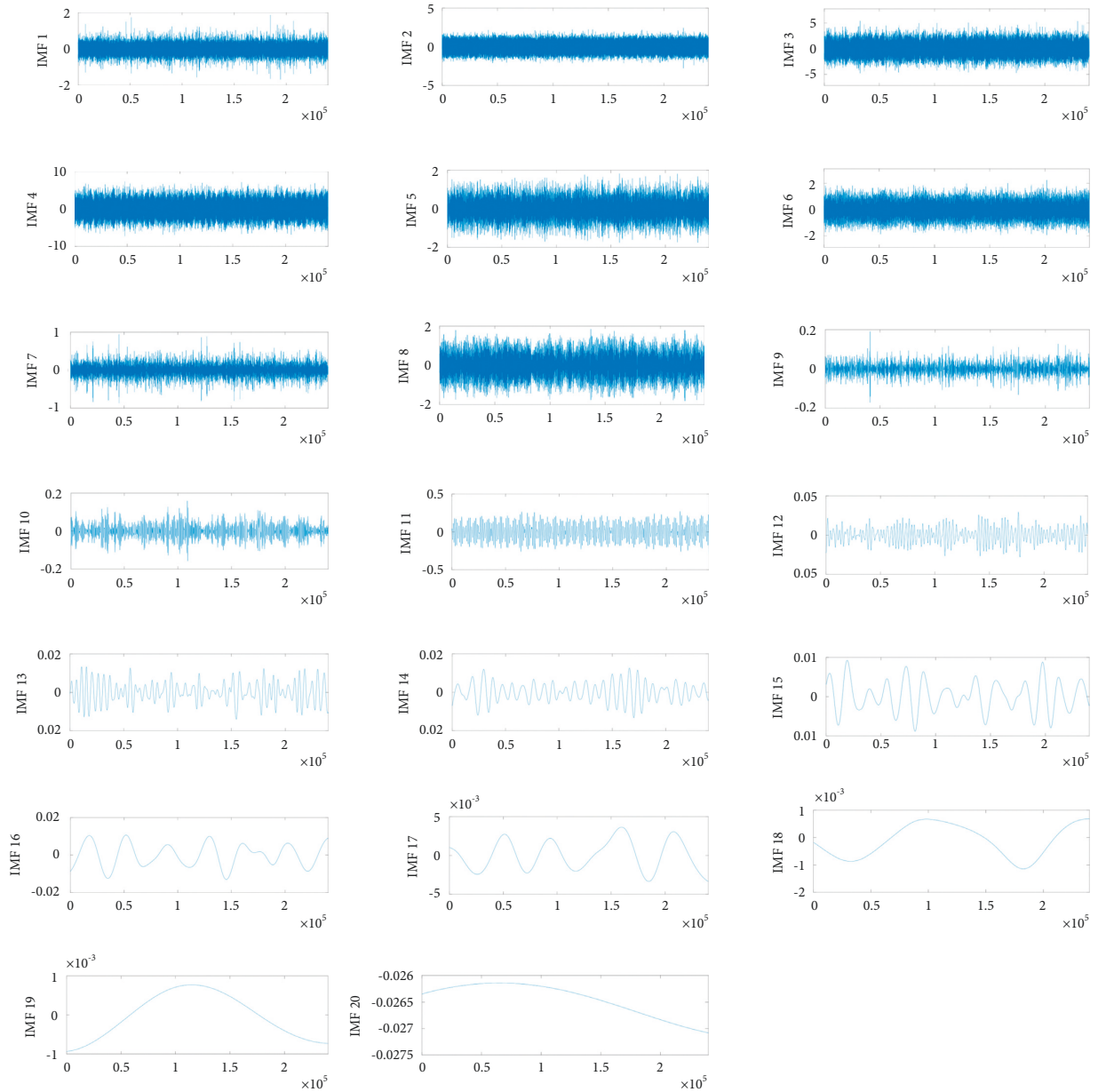


FIGURE 6: Decomposition results by using NA-MEMD on the synthetic multivariate signal.

nodes in the output layer is considered to be 2, corresponding to two conditions of system. 100 feature vectors are constructed for each system condition. 80% of data are considered as the training data and 20% as the testing data. In this study, a three-layer neural network is constructed for the intelligent fault diagnosis procedure. Therefore, according to (17) and network training conditions, the number of hidden layer nodes (k) is 18–26. In this study, $k = 18$, since the differences between the output errors for the different values of k are in the same order

($1e-5$) while the learning rate is 0.001. The training function is TRAINLM to update the weight and bias values based on the Levenberg–Marquardt optimization method, and the activation function between the hidden layer and output layer is Sigmoid function. Note that the setting is similar for all experiments and all experiments have been done using MATLAB platform. Diagnosis rate for both training data and test data is 100%. This shows that the features and the network configuration are successfully selected.

TABLE 6: Discrimination criterion for the proposed features.

Feature	AF_6	AF_8	E_3'	E_4'	pt_1	E_5'	E_6'	E_7'	AF_{11}	E_8'	$pt_i, pf_{1i}, pf_{i'}^*$
J criterion's value	1.932	1.812	1.720	1.600	1.541	1.021	0.952	0.741	0.603	0.402	≤ 0.1

TABLE 7: Neural network input and output vector.

Feature vector	System condition
[0.0826, 0.7469, 0.0877, 0.0200, 0.0261, 0.0367, 0.0477, 0.0409, 0.0535]	Healthy
[0.2125, 0.5985, 0.0347, 0.0664, 0.0052, 0.0826, 0.5308, 0.7147, 0.0659]	Faulty

6. Conclusion

In this paper, the MEMD algorithm is applied for extracting features from rotating machinery. To investigate the capacity of the proposed method, vibration signals from a wind turbine gearbox system as a rotating machinery system are utilized. When the rotating system is complex and consists of many faults, multiple sensors are exploited to obtain comprehensive information from the system. MEMD algorithm has the advantage of dealing with multivariate signals simultaneously. Usually, when the system is nonstationary and there are nonlinearity and multiple faults, using traditional features may be abortive. Features derived from the MEMD algorithm are based on the time and frequency domain, which compensate for the problem of using traditional features. To validate the effectiveness of the proposed features, a discrimination criterion is introduced. This criterion is based on the relativity of features to the fault classes.

The basic MEMD algorithm is sensitive to noise. In this study, an extension of MEMD called NA MEMD is implemented on multivariate signals to overcome the noise sensitivity of MEMD. MEMD algorithm decomposes signals into some signals named IMFs. Some of these IMFs are spurious and need to be eliminated from the calculation. A correlation factor is introduced to achieve this purpose. With the help of this factor, the number of redundant features is reduced. Two types of features are extracted from the IMFs. From the point of view of time-domain analysis, the energy moment of IMFs is a suitable feature, since it contains the time characteristics of signals. Therefore, this can be helpful when the signal is nonstationary. The other feature is in the frequency domain, and it relates to the amplitude of frequency spectrum in the characteristic frequencies. Because each IMF order encompasses a small frequency range, frequency analysis of IMFs is an effective way of highlighting characteristics.

Based on the results, designing a neural network using the proposed features yields acceptable output results. The network is successfully trained using the training data, and the diagnostic rate is 100% not only for the training data, but also for the test data. It should be mentioned that the proposed algorithm is applied to real experimental data; however, by increasing the number of classes, the performance may decrease.

It should be noted that intelligent feature extraction using the proposed NA-MEMD method provides comprehensive information on the health status of the system. The

proposed methodology gives higher explainability of the features compared to other similar methods. However, recently, deep learning-based methods have been successfully implemented in industrial datasets to automatically extract features. In spite of the effectiveness of these methods, they require high computation resources compared with the proposed method.

Data Availability

The NREL wind turbine data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was funded by the Key R&D Plan of China for Winter Olympics (2021YFF0306401) and the Key Special Project of the National Key Research and Development Program "Technical Winter Olympics" (2018YFF0300502 and 2021YFF0306400).

References

- [1] D. Gorecky, M. Schmitt, M. Loskyll, and D. Zühlke, "Human-machine-interaction in the industry 4.0 era," in *Proceedings of the 2014 12th IEEE International Conference on Industrial Informatics*, pp. 289–294, Porto Alegre, Brazil, 2014.
- [2] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Transfer learning using deep representation regularization in remaining useful life prediction across operating conditions," *Reliability Engineering & System Safety*, vol. 211, p. 107556, 2021.
- [3] M. Mousavi, M. Alzgoool, and S. Towfighian, "Autonomous shock sensing using bi-stable triboelectric generators and mems electrostatic levitation actuators," *Smart Materials and Structures*, vol. 30, no. 6, Article ID 065019, 2021.
- [4] W. Zhang and X. Li, *Federated Transfer Learning for Intelligent Fault Diagnostics Using Deep Adversarial Networks with Data Privacy*, IEEE/ASME Transactions on Mechatronics, 2021.
- [5] I. Jebellat, H. N. Pishkenari, and E. Jebellat, "Training microrobots via reinforcement learning and a novel coding method," in *Proceedings of the 2021 9th RSI International Conference on Robotics and Mechatronics (ICRoM)*, pp. 105–111, IEEE, Tehran, Iran, 2021.

- [6] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Open-set domain adaptation in machinery fault diagnostics using instance-level weighted adversarial learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 11, pp. 7445–7455, 2021.
- [7] S. Siahpour, X. Li, and J. Lee, "A novel transfer learning approach in remaining useful life prediction for incomplete dataset," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, 2022.
- [8] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, *Degradation Alignment in Remaining Useful Life Prediction Using Deep Cycle-Consistent Learning*, IEEE Transactions on Neural Networks and Learning Systems, 2021.
- [9] V. Fazlollahi, F. A. Shirazi, M. Taghizadeh, and S. Siahpour, "Robust wake steering control design in a wind farm for power optimisation using adaptive learning game theory (algt) method," *International Journal of Control*, pp. 1–17, 2021.
- [10] A. Ainapure, S. Siahpour, X. Li, F. Majid, and J. Lee, "Intelligent robust cross-domain fault diagnostic method for rotating machines using noisy condition labels," *Mathematics*, vol. 10, no. 3, p. 455, 2022.
- [11] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Universal domain adaptation in fault diagnostics with hybrid weighted deep adversarial learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 7957–7967, 2021.
- [12] S. Siahpour, F. N. Khakiani, V. Fazlollahi, A. Golozar, and F. A. Shirazi, "Morphing omni-directional panel mechanism: a novel active roof design for improving the performance of the wind delivery system," *Energy*, vol. 217, p. 119400, 2021.
- [13] C. Shen, D. Wang, F. Kong, and P. W. Tse, "Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier," *Measurement*, vol. 46, no. 4, pp. 1551–1564, 2013.
- [14] B. Li and Y. Zhang, "Supervised locally linear embedding projection (sllep) for machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 25, no. 8, pp. 3125–3134, 2011.
- [15] Y. Lei, Z. He, and Y. Zi, "A new approach to intelligent fault diagnosis of rotating machinery," *Expert Systems with Applications*, vol. 35, no. 4, pp. 1593–1600, 2008.
- [16] J. Ben Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, vol. 89, pp. 16–27, 2015.
- [17] H. D. M. de Azevedo, P. H. C. de Arruda Filho, A. M. Araújo, N. Bouchonneau, J. S. Rohatgi, and R. M. C. de Souza, "Vibration monitoring, fault detection, and bearings replacement of a real wind turbine," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 39, no. 10, pp. 3837–3848, 2017.
- [18] T. Barszcz and R. B. Randall, "Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine," *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1352–1365, 2009.
- [19] J. Igba, K. Alemzadeh, C. Durugbo, and E. T. Eiriksson, "Analysing rms and peak values of vibration signals for condition monitoring of wind turbine gearboxes," *Renewable Energy*, vol. 91, pp. 90–106, 2016.
- [20] M. Inalpolat and A. Kahraman, "A theoretical and experimental investigation of modulation sidebands of planetary gear sets," *Journal of Sound and Vibration*, vol. 323, no. 3–5, pp. 677–696, 2009.
- [21] M. E. Badaoui, F. Guillet, and J. Danière, "New applications of the real cepstrum to gear signals, including definition of a robust fault indicator," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1031–1046, 2004.
- [22] C. J. Li and S. Wu, "On-line detection of localized defects in bearings by pattern recognition analysis," *Journal of Manufacturing Science and Engineering disseminates*, 1989.
- [23] L. Satish, "Short-time fourier and wavelet transforms for fault detection in power transformers during impulse tests," *IEE Proceedings - Science, Measurement and Technology*, vol. 145, no. 2, pp. 77–84, 1998.
- [24] J. Lin and L. Qu, "Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis," *Journal of Sound and Vibration*, vol. 234, no. 1, pp. 135–148, 2000.
- [25] X. Fan and M. J. Zuo, "Machine fault feature extraction based on intrinsic mode functions," *Measurement Science and Technology*, vol. 19, no. 4, Article ID 045105, 2008.
- [26] W. J. Staszewski, K. Worden, and G. R. Tomlinson, "Time-frequency analysis in gearbox fault detection using the wigner-ville distribution and pattern recognition," *Mechanical Systems and Signal Processing*, vol. 11, no. 5, pp. 673–692, 1997.
- [27] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903–995, 1998.
- [28] D. Yu, J. Cheng, and Y. Yang, "Application of emd method and hilbert spectrum to the fault diagnosis of roller bearings," *Mechanical Systems and Signal Processing*, vol. 19, no. 2, pp. 259–270, 2005.
- [29] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, 2004.
- [30] X. Fan and M. J. Zuo, "Gearbox fault detection using empirical mode decomposition," *ASME International Mechanical Engineering Congress and Exposition*, vol. 47160, pp. 37–45, 2004.
- [31] I. Antoniadou, G. Manson, W. J. Staszewski, T. Barszcz, and K. Worden, "A time-frequency analysis approach for condition monitoring of a wind turbine gearbox under varying load conditions," *Mechanical Systems and Signal Processing*, vol. 64–65, pp. 188–216, 2015.
- [32] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 01, no. 01, pp. 1–41, 2009.
- [33] Y. Lv, R. Yuan, and G. Song, "Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing," *Mechanical Systems and Signal Processing*, vol. 81, pp. 219–234, 2016.
- [34] D. Looney and D. P. Mandic, "Multiscale image fusion using complex extensions of emd," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1626–1630, 2009.
- [35] G. Rilling, P. Flandrin, P. Goncalves, and J. M. Lilly, "Bivariate empirical mode decomposition," *IEEE Signal Processing Letters*, vol. 14, no. 12, pp. 936–939, 2007.
- [36] N. ur Rehman and D. P. Mandic, "Empirical mode decomposition for trivariate signals," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1059–1068, 2009.
- [37] N. Rehman and D. P. Mandic, "Multivariate empirical mode decomposition," *Proceedings of the Royal Society A: Mathematical, Physical & Engineering Sciences*, vol. 466, no. 2117, pp. 1291–1302, 2010.

- [38] X. Zhao, T. H. Patel, and M. J. Zuo, "Multivariate emd and full spectrum based condition monitoring for rotating machinery," *Mechanical Systems and Signal Processing*, vol. 27, pp. 712–728, 2012.
- [39] R. Ricci and P. Pennacchi, "Diagnostics of gear faults based on emd and automatic selection of intrinsic mode functions," *Mechanical Systems and Signal Processing*, vol. 25, no. 3, pp. 821–838, 2011.
- [40] S. Siahpour, X. Li, and J. Lee, "Deep learning-based cross-sensor domain adaptation for fault diagnosis of electro-mechanical actuators," *International Journal of Dynamics and Control*, vol. 8, no. 4, pp. 1054–1062, 2020.
- [41] V. Vakharia, V. K. Gupta, and P. K. Kankar, "Efficient fault diagnosis of ball bearing using relieff and random forest classifier," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 39, no. 8, pp. 2969–2982, 2017.
- [42] Y. Yu, Y. Dejie, and C. Junsheng, "A roller bearing fault diagnosis method based on emd energy entropy and ann," *Journal of Sound and Vibration*, vol. 294, no. 1-2, pp. 269–277, 2006.
- [43] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets-Empirical mode decomposition feature extraction and neural network," *Mechanical Systems and Signal Processing*, vol. 27, pp. 696–711, 2012.
- [44] N. ur Rehman, C. Park, N. E. Huang, and D. P. Mandic, "Emd via memd: multivariate noise-aided computation of standard emd," *Advances in Adaptive Data Analysis*, vol. 05, no. 02, Article ID 1350007, 2013.
- [45] G. Qu, S. Hariri, and M. Yousif, "A new dependency and correlation analysis for features," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 9, pp. 1199–1207, 2005.
- [46] L. Tlig, M. Sayadi, and F. Fnaiech, "A new fuzzy segmentation approach based on s-fcm type 2 using lbp-gco features," *Signal Processing: Image Communication*, vol. 27, no. 6, pp. 694–708, 2012.
- [47] H. Yuan, X. Wang, X. Sun, and Z. Ju, "Compressive sensing-based feature extraction for bearing fault diagnosis using a heuristic neural network," *Measurement Science and Technology*, vol. 28, no. 6, Article ID 065018, 2017.
- [48] S. Sheng, *Wind Turbine Gearbox Condition Monitoring Round Robin Study-Vibration Analysis*, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2012.
- [49] H. H. Mabie and C. F. Reinholtz, *Mechanisms and Dynamics of Machinery*, John Wiley & Sons, 1991.
- [50] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics-A tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, 2011.

Research Article

Sign Language Recognition for Arabic Alphabets Using Transfer Learning Technique

Mohammed Zakariah ^{1,2} **Yousef Ajmi Alotaibi** ^{1,2} **Deepika Koundal**³ **Yanhui Guo**⁴ and **Mohammad Mamun Elahi** ⁵

¹College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

²Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, P.O. Box 57168, Riyadh 21574, Saudi Arabia

³Department of Systemics, University of Petroleum & Energy Studies, Dehradun, India

⁴University of Illinois, Springfield, USA

⁵Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh

Correspondence should be addressed to Mohammad Mamun Elahi; mmelahi@cse.uui.ac.bd

Received 5 January 2022; Revised 24 January 2022; Accepted 5 April 2022; Published 22 April 2022

Academic Editor: Zhongxu Hu

Copyright © 2022 Mohammed Zakariah et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sign language is essential for deaf and mute people to communicate with normal people and themselves. As ordinary people tend to ignore the importance of sign language, which is the mere source of communication for the deaf and the mute communities. These people are facing significant downfalls in their lives because of these disabilities or impairments leading to unemployment, severe depression, and several other symptoms. One of the services they are using for communication is the sign language interpreters. But hiring these interpreters is very costly, and therefore, a cheap solution is required for resolving this issue. Therefore, a system has been developed that will use the visual hand dataset based on an Arabic Sign Language and interpret this visual data in textual information. The dataset used consists of 54049 images of Arabic sign language alphabets consisting of 1500\ images per class, and each class represents a different meaning by its hand gesture or sign. Various preprocessing and data augmentation techniques have been applied to the images. The experiments have been performed using various pretrained models on the given dataset. Most of them performed pretty normally and in the final stage, the EfficientNetB4 model has been considered the best fit for the case. Considering the complexity of the dataset, models other than EfficientNetB4 do not perform well due to their lightweight architecture. EfficientNetB4 is a heavy-weight architecture that possesses more complexities comparatively. The best model is exposed with a training accuracy of 98 percent and a testing accuracy of 95 percent.

1. Introduction

Over 70 million people use sign language worldwide and an automated process for interpreting it might significantly impact communication between those who use it and those who do not. Sign language is a kind of non-verbal communication that includes other bodily organs. In sign language communication, facial expressions, eye, hand, and lip gestures are used to transmit data. Individuals who are deaf or hard of hearing rely heavily on sign language as a form of communication in their daily lives [1].

As per the World Health Organization, hearing impairment affects 5% of the Earth's population. However, this appears to be a minor figure, it indicates that hearing impairment affects over 460 million individuals worldwide; 34 million of whom are youngsters. Moreover, it is predicted that even by 2050, over 900 million individuals will undergo hearing impairment [2], with 1.1 billion youth at risk of deafness due to noise exposure and other difficulties. Untreated hearing loss costs the world 750 billion US dollars [2]. Based on the severity of the deafness, hearing impairment is classified as mild, moderate, severe, or profound. People with severe or profound hearing impairment cannot

attend to others and consequently have communication difficulties. This poor communication could significantly influence the deaf person's mental health, including loneliness, solitude, and dissatisfaction. The deaf society communicates using a gesture-based language known as sign language. Deaf individuals use sign language motions to connect. On the other hand, the hearing society does not recognize these gestures, which creates a communication barrier between a deaf and a hearing individual. There are almost 200 sign languages globally, and sign languages, like spoken languages, vary from each other.

Sign language is a subset of communication used as a medium of interaction by the deaf. Unlike other natural languages, it uses significant bodily motions to communicate messages, known as gestures or signs. To communicate a message, hand and finger gestures, head nodding, shoulder gestures, and facial expressions are employed. Therefore, the proposed work would help deaf people to interact between deaf and deaf or deaf and normal persons. When a deaf or hard-of-hearing person tries to express anything, they use gestures to communicate. Each symbol represents a different letter, word, or emotion. A phrase is formed by the combination of signals, much as the string of words includes words in spoken languages. As a result, sign language is a fully formed natural language with grammar and sentence structure.

Humans need verbal communication to carry out social tasks. Consequently, voiceless or silent (D&M) persons were also incapable of conversing vocally with others. Those who communicate through sign language [3] can overcome this issue. The visual modality to express meaning is known as sign language. The message or feel is represented through a manual sign sequence in conjunction, such as nonmanual elements of the communication. Forms of communication vary from one another and are not mutually exhaustive [4]. Sign languages have their own rules and components, respectively, manual and nonmanual [5, 6]. American Sign Language (ASL), British Sign Language (BSL), Brazilian Sign Language (LIBRAS), Japanese Sign Language (JSL), Arabic Sign Language (ArSL), Hindustan Sign Language (ISL), and Bangla Sign Language (BdSL) are some of the sign languages used across the globe [7]. Sign languages are frequently not understood by those who can talk and hear. Written language plays a minor part in establishing communication between D&M societies and the wider public, as much D&M lacks proficiency in spoken language. Again, this technique is hugely sluggish in immediate and emergency face-to-face conversations [8]. According to reports, over 16 million individuals in Bangladesh are deaf, deafened, or have auditory impairments [9]. They adopt sign language, which most people cannot recognize, to describe their emotions. Interaction between D&M personnel and the general public necessitates translating sign language into a language that the general public can identify.

Deep learning remained a class of learning algorithms developed to describe complex structures by combining numerous nonlinear adjustments. The neural networks linked to building deep neural networks are the essential building blocks of deep learning. These methods have

enabled significant progress in sound and picture processing, encompassing face identification, computer vision, voice recognition, automated language processing, text categorization (spam identification), and a diversity of other fields like drug diagnosis and genomics. There are several potential uses. First, deep learning enables computational algorithms with several processing layers to acquire a representation of different abstracted dimensions. Deep learning detects unpredictability in large datasets by using the backpropagation technique to express how a system should modify its inner parameters, which have been used to perform a presentation in each level from the symbolization in the preceding layer. Whereas recurrent networks have cast a flashlight on sequential information, such as voice and text, Deep Convolution Network (DCN) has made significant advances in processing video, picture, audio, and speech. Third, deep learning is often carried out using neural network building. The term "deep" mentions the total number of layers in a network; the more layers, the deeper the system. Third, deep learning is extraordinary in terms of precision. Modern tools and tactics have greatly improved deep learning algorithms to the point where they can outperform human performance. This degree of accuracy is made possible by three innovation-enabling influencing factors: The main aim is to develop a sign language recognition system capable enough of translating the most commonly expressed hand gestures used by deaf or dumb people into textual data. To make these disabled people communicable is the prime objective. The contributions are listed as follows: (i) Several data preprocessing techniques have been applied to make the training process faster and less complex to simplify the model training and evaluation process. (ii) Transformation of inconsistent and irregular Arabic datasets has been done into the proper format by various data augmentation techniques. (iii) The proposed work is based on transfer learning using several architectures pretrained on the ImageNet dataset. Those architectures are customized to make them adaptable for the current problem domain. (iv) The experimental work was carried out to test the pretrained models on the unforeseen data. (v) Several Keras pretrained models have been adopted and convolutional neural network architectures are applied for the given case in which the EfficientNetB4 model has outperformed all other models.

The rest of the paper is organized as follows: Section 2 discusses the related work and different techniques applied in this domain. Section 3 presents the material and methodology of the proposed work. Subsequently, results and discussion have been presented in Sections 4 and 5, respectively. Finally, the paper is concluded with a conclusion in Section 6.

2. Related Work

Arabic is the world's 4th most spoken language (Generates a set Consulting Group 2020). Arabic Sign Language (ArSL) seems to be the certified primary language again for talking and listening impaired in Arab countries. The Arab Federation of the Deaf publicly established this in 2001. Even

though Arabic is among the world's main languages, ArSL is still in its early stages. The most common problem that ArSL patients face is "diglossia." Regional dialects are spoken rather than written languages across every country. As a result, various spoken dialects produced varied ArSLs. They are as abundant as Arab states, yet they share several terms and an alphabet. "ArSL is dependent on the alphabet." Arabic is a sophisticated and pleasant language and one of the Semitic languages vocalized by about 380 million individuals worldwide as their primary official language. Arabs demonstrate plausible semantic and intellectual unity [10].

The authors in this work [11] concentrated on NN's ability to aid with ArSL hand gesture identification. The purpose of the study was to show the use of several types of NN through living person gesture recognition, including stationary and dynamic indicators. First, they demonstrated the practice of Feed Forward Neural Network (FFNN) and RNN in conjunction with its different topologies, completely and moderately reoccurring systems. They then examined their offered structure; the evaluation results revealed that the suggested form with the entire repeated design does have an implementation with a precision rate of 95% for stationary action recognition.

In this study [12], the authors emphasized the automated acknowledgment of the ArSL alphabets using a picture-based method. In particular, several visual features were investigated to construct an accurate ArSL alphabets sensor. One-Versus-All SVM received the extracted visible tags. The results revealed that the Histogram of Oriented Gradients (HOG) signifier outruns other characteristics. As a result, the ArSL gestures system trained by One-Versus-All SVM using HOG identifiers was developed in this study. The authors in this work [13] used the Kinect Sensor to make a Real-Time System for automatic ArSL identification structure based on the Dynamic Time Warping coordination method. The program does not use any power/data gloves. Many trials were used to detect for a lexicon of 30 distinct phrases specifically produced signals again from standardized ArSL. The architecture could function in three means: digitally, signer-independent, and signer-dependent. They used the Dynamic Time Warping coordination method to differentiate between indications. The tests showed that the current version has a high detection score for each option. The framework achieved a detection accuracy of 97.58 percent and a ratio of error of 2.42 percent for signer-dependent. The algorithm then achieved a detection accuracy of 95.25 percent and a ratio of error of 4.75 percent for signer-independent recognition. In some other works conducted by [14, 15], various aspects of human-computer interactions were discussed.

Alternative techniques to sign lingual identification are focused on Hidden Markov Models, like studies from 2011 that identify Arabic Sign Language including the efficiency of up to 82.22 percent [16]. Some other studies that used Hidden Markov Models can be found in [17]. At the same time, in [18], a five-stage procedure for an Arabic sign language translator was published, concentrating on background subtraction of transcription, magnitude, or partially invariant, and achieving an efficiency of 91.3 percent.

Almasre and Al-Nuaim employed unique detectors like the Microsoft Kinect or Leap Motion Detectors for record-keeping throughout one's hand-gesturing system to identify 28 Arabic Sign Language motions [19]. Recent work upon Arabic sign language identification has been revealed throughout [20]. Many CNNs have been formed and offered input from an imaging system that contained the elevation and breadth of items and their intensity. The figures are instead processed by a CNN based on the frame rate of the depth footage, which also determines how extensive the system is. Lower frame rates result in less depth, whereas faster refresh rates result in further detail.

In this work [21], a novel model was introduced for Arabic Sign Language Acknowledgment in 2019 utilizing Convolutional Neural Network (CNN) to recognize 28 Arabic letters and numerals ranging from 0 to 10 from an image dataset of 7869 pictures. The suggested framework had seven layers and was instructed numerous times on various training-testing variations, with the highest correctness seeming to be 90.02 percent with a picture training data of 80%. Eventually, the researchers contrasted with other methods, demonstrating the suggested model's benefit. CNN is a deep neural network category that is most widely used in computer field vision. Vision-based techniques primarily concentrate on acquired pictures of the motion and extract the principal characteristic to recognize it. This technology has been used to solve a variety of problems involving superresolution, picture segmentation and semantic breakdown, multimedia systems, and emotion identification [22–24]. In a similar effort [25], Oyedotun and Khashman were among the few well-known scholars that employed CNN in conjunction with Stacked Denoising Autoencoder (SDAE) to recognize 24 hand motions in American Sign Language (ASL) obtained from a communal record. On the other hand, Pigou et al. proposed using Convolutional Neural Network (CNN) to identify Italian sign language [8]. However, Hu et al. had developed a suggestion for the design of hybrid CNN and RNN to preserve the temporal features correctly for the electromyogram signal, which addresses the issue of action identification. Another work [26] describes an extraordinary CNN model that automatically detects numbers relying on hand signals and communicates the specific outcome in the Bangla language, which is followed in this study. In a similar work [27], a CRNN module for hand pose estimation is conducted. There is also a suggestion in [28] to employ transfer learning on data acquired from many individuals, simultaneously utilizing a deep-learning system to understand discriminant traits discovered in massive datasets. A deep convolutional neural network-based Bernoulli heatmap for head pose estimation was conducted by [29]. Another work [30] related to 3D separable convolutional neural network for dynamic hand gesture recognition is used for recognizing the hand gesture. Another work [31] applied flexible strain sensors for wearable hand gesture recognition, which is the latest in this field of research. Further to the latest work-related hand gesture, the authors here [32] have applied deformable convolution neural networks. Fingerprint detection [33] for the recognition of hand gestures is



FIGURE 1: Dataset overview.

another latest work proposed for HCI. A lightweight neural network [34] is applied for hand gesture recognition. Geometric features learning [35] is another technique to recognize hand gestures. In [36], a consistent identification system is suggested employing the K -nearest neighbor classifier and statistical feature extraction approach for the Arabic sign language as another methodology for recognizing the Arabic sign language. Sadly, the fundamental disadvantage of Tubaiiz's technique is that consumers are forced to utilize instrumented hand gloves to collect the specific gesture's details, which frequently creates excellent suffering to the consumer. Following that, [37] suggests developing an instrumented glove to create an Arabic sign language recognition system. They presented constant detection of Arabic sign language employing hidden Markov models and spatiotemporal characteristics. Hand pose estimation with a multiscale network was proposed by [38]. Similarly, [39] studied translation from Arabic sign language to text, which may be utilized on portable devices. The automated identification of Arabic sign language utilizing sensor and image techniques is reported in [40]. In [41], using two depth sensors to identify Arabic Sign Language (ArSL) hand movements proposes a flexible Arabic Sign Language identification structure based on two machine learning algorithms that use Microsoft Kinect. Furthermore, the current CNN technique to Arabic sign language has been unparalleled in the sign language study arena [42]. As a result, the objective of this study is to build a vision-based organization that recognizes Arabic hand sign-based letters and converts them into Arabic language using CNN. For each of the 31 letters of Arabic sign language, a collection of 100 photos in the training set and 25 pictures in the test set is

constructed. Several hyperparameter combinations evaluate the proposed system to get the best outcomes with the lowest amount of training duration.

3. Material and Methods

3.1. Dataset. The dataset consists of 54049 images of Arabic sign language alphabets performed by more than 40 people for 32 standard Arabic signs and alphabets. The dataset is available at ArSL2018 [43], launched by Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia, to be open to Machine Learning and Deep Learning researchers. The number of images per class differs from one type to another. Each distinct hand gesture indicates some meaningful information. There are around 1500 images per class, and each class represents a different meaning by its hand gesture or sign. Pictorially, the sample image of each class and its label is illustrated in Figure 1.

For some storage schemes, 32 folders are created, and each folder consists of around 1500 images incorporating differently aged people's hand gestures in different environments. The directories containing these folders are treated as training and validation datasets for the model, which will be explained later in this section. Before talking about the model used, it is mandatory to undergo data preprocessing to make the dataset more consistent and compatible with the model as an input. So, how the data preprocessing is done is elaborated in the next section.

3.2. Methodology. Before talking about the model used, it is mandatory to undergo data preprocessing to make the

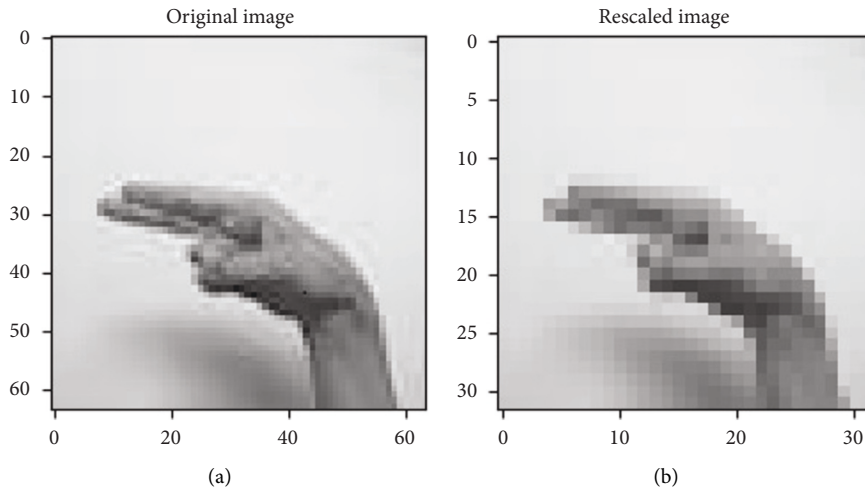


FIGURE 2: (a) Original image versus (b) rescaled image.

dataset more consistent and compatible with the model as an input. So, how the data preprocessing is done is elaborated in the next section.

3.2.1. Data Preprocessing. The data preprocessing involves the transformation applied to the data before feeding it to the model for training/testing. So, what changes are performed on the dataset is described below. As already mentioned, the number of images per class differs. This imbalance meant among the classes may degrade the training performance of the model. Thus, there must be an equal number of images among all classes to avoid this imbalance meant. This imbalance is removed by looping over each class folder to get the filenames of all the images per class. 1000 images are picked randomly from the current class folder during each iteration, and the rest are removed. Resultantly, 32000 images are filtered by summing up 1000 images of all the classes. The images contained in each class have the dimensions of (64×64) . To keep the computations while training less complex and fast, the images can be rescaled into (32×32) following the same dimensionality ratio. Rescaling is represented pictorially in Figure 2.

(1) Data Augmentation. The data augmentation technique is widely used to increase the size of the training dataset by generating artificial modified versions of the original images from the training dataset. The technique results in a more diverse and consistent sequence of images, further creating more generalized and skillful deep learning models. The technique helps avoid overfitting and underfitting the model by applying several optional modifications to the training images. In this case, the following augmented changes are performed on the training images through ImageDataGenerator provided by the Keras API [44]. This augmentation technique includes the horizontal shifting of the object to the left or right up to the defined limit, as shown in Figure 3.

This step includes the vertical shifting of the targeted object to up and down up to a certain limit, as shown in Figure 4. This augmentation technique involves the random

darkening and brightening the images up to a certain limit, as shown in Figure 5. This augmentation technique randomly removes or adds the pixels into the images for zoom in or zoom out up to the provided limitation, as shown in Figure 6.

All the above-mentioned augmentation techniques are performed by passing parameters with their limitations to the ImageDataGenerator class provided by Keras API. The transformations of the original image can be seen in Figure 7. It includes various augmented images generated from the one original image belonging to class “khaa.” These images are then converted into normalized images, and this normalization process is explained in the next section. The data normalization step performs the normalization process on each image of the dataset. Usually, the pixel values in the image range from 0 to 255. But these values must be rescaled before providing these images to the model as an input. So, the normalization will rescale these pixel values in the range of $(0, 1)$. This rescaling will keep the model easy to learn and train fast, and this is represented in Figure 8.

Considering Figure 8, there is some contrast difference between the two images. The normalized image is more precise and brighter than the original image. So, normalized images are more adaptable and easier for the model to train.

(2) Data Splitting. The data used to build the model comes from multiple types of datasets. There are three different purposeful datasets for any computer vision project to analyze, compare, and improve the model’s performance. In particular, these three different types of datasets are used in various stages of creating any machine learning model. These three distinct datasets are stated below:

Training dataset on which the model is trained for learning weights or features. Initially, the model is fitted on the training dataset, and in our case specifically, 80 percent of the whole dataset is used for the training dataset, which is approximately 25600 images. Validation dataset the model is fitted on this dataset for the unbiased evaluation of itself during training. It validates the model’s performance based on how well the model learns its weights before it is used for real-time testing on the testing dataset. In our scenario of

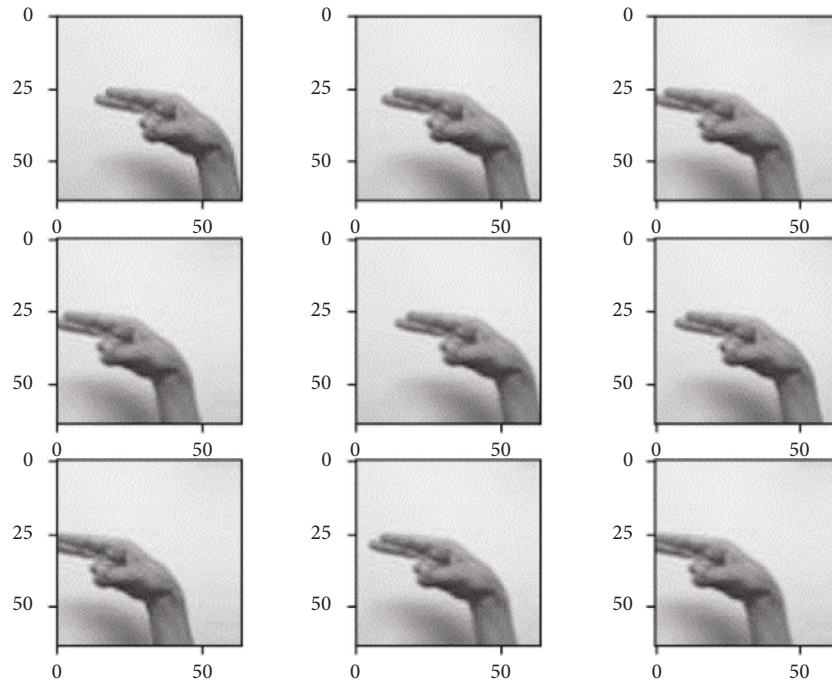


FIGURE 3: Width-shift augmented images.

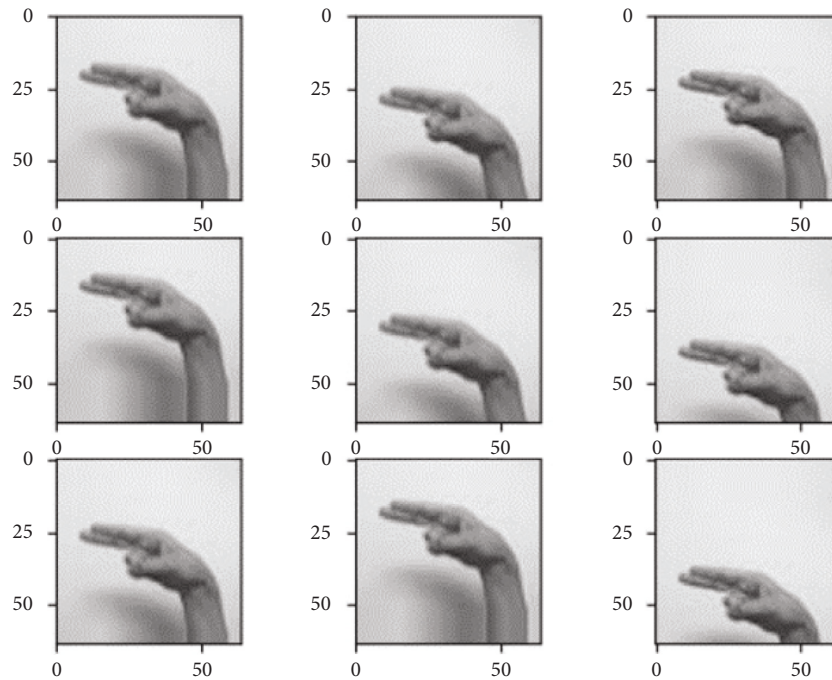


FIGURE 4: Height-shifted augmented images.

sign language recognition, 20 percent of the dataset is used, which is equivalent to 6400 images. Test dataset after the completion of training and validation phenomena is used to examine the performance of the proposed model and measure its efficiency and accuracy and how well the model is trained. Nine hundred sixty samples are used for the test dataset since there are 30 test images for each sign alphabet. This self-generated test set is created to measure the model's

ability to generalize. More importantly, this test set is not collected from the 32000 images.

After the dataset is fully preprocessed, it is fed to the model network in a compatible input fashion for training. But to start the training, it is vital for the reader to understand the workflow, as shown in Figure 9.

Before starting this time-consuming process, it is necessary to ensure the best possible selection of the

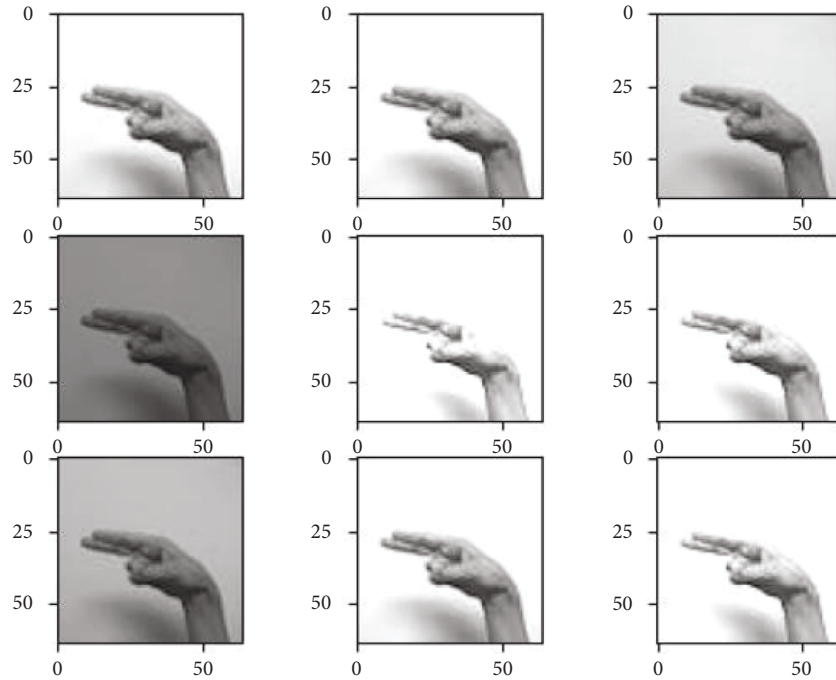


FIGURE 5: Brightness-ranged augmented images.

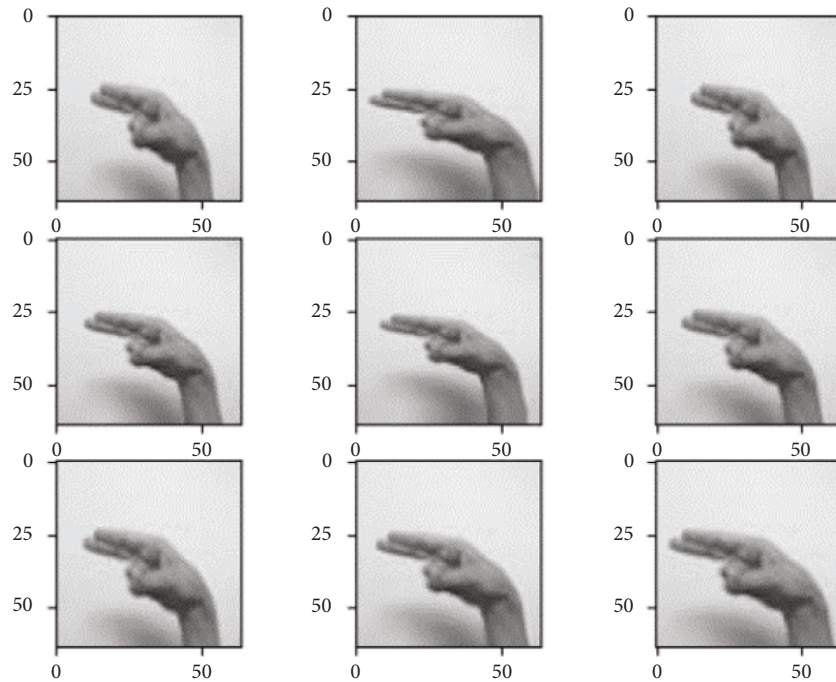


FIGURE 6: Zoom-range augmented images.

deep neural network considering the problem domain. Various frameworks can be used in this case, like TensorFlow, Keras, PyTorch, and so on. Each framework has its pros and cons; considering the problem domain, Keras is used. So, to ensure the best possible fit, there are several pretrained models available in the Keras library.

Those pretrained models are trained at the ImageNet dataset to provide state-of-the-art results in the domain of image classification. So, here the question arises that what is the ImageNet dataset and what classes the ImageNet dataset constitutes are explained briefly in the next section.

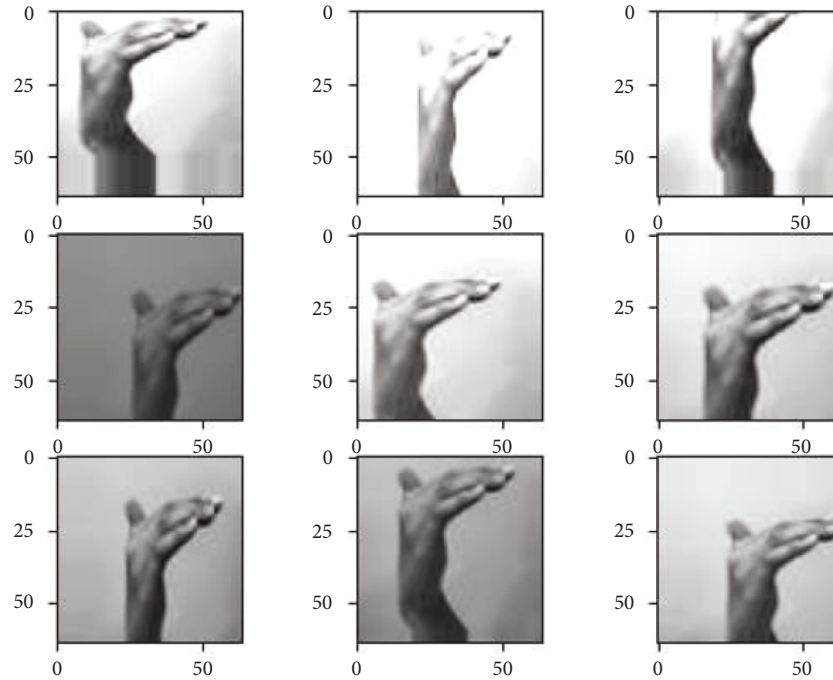


FIGURE 7: . Augmented images generated from the original image.

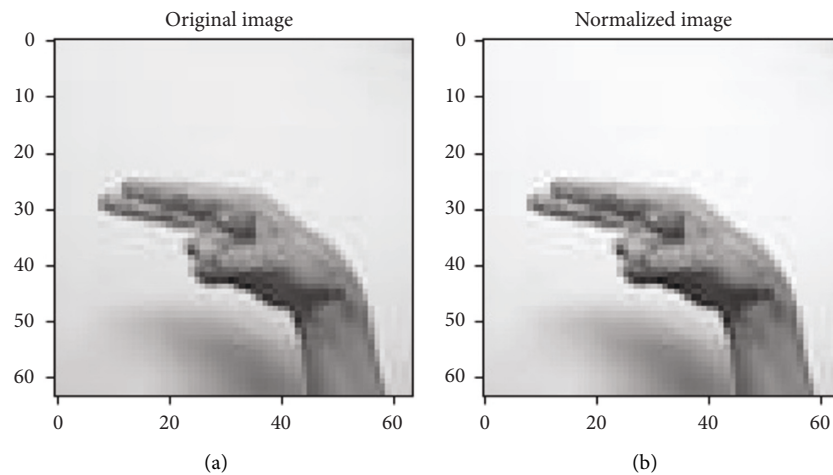


FIGURE 8: (a) Original image versus (b) normalized image.

3.2.2. Keras Pretrained Models. ImageNet is an extensive collection of annotated images publicly made available for computer vision research. This large-scale collection of images is a critical resource for analyzing, training, and testing the machine learning algorithms. There are around 14 million images, and 1000 categories or classes in this dataset, and this dataset is also used for large-scale visual recognition challenge competitions. The pretrained models provided by the Keras Applications' Python package is also complex functional models because these applicable models are trained on the ImageNet dataset. These pretrained models can classify any image that falls into these categories of images.

As mentioned before briefly, Keras applications constitute several pretrained deep learning models available in its repository. The pretrained weights are also available

alongside these models. So, these models are further used for custom object detection, image classification, and so on. But considering the domain problem, image classifiers are filtered from these pretrained models and not the object detectors because the case requires performing the image classification. The selection is made considering the hand gesture dataset's complexity and nature. The selected pretrained models with their results are mentioned in Table 1.

In Table 1, it is essential to note that all the models are trained using the ImageNet dataset. Every model has its size, accuracy, and several parameters along with the architecture depth. These models are retrained further on the Arabic hand gesture classification dataset comprising 32 classes. After training, the best possible fit is considered for the case. So, the final selection is made after custom-training these

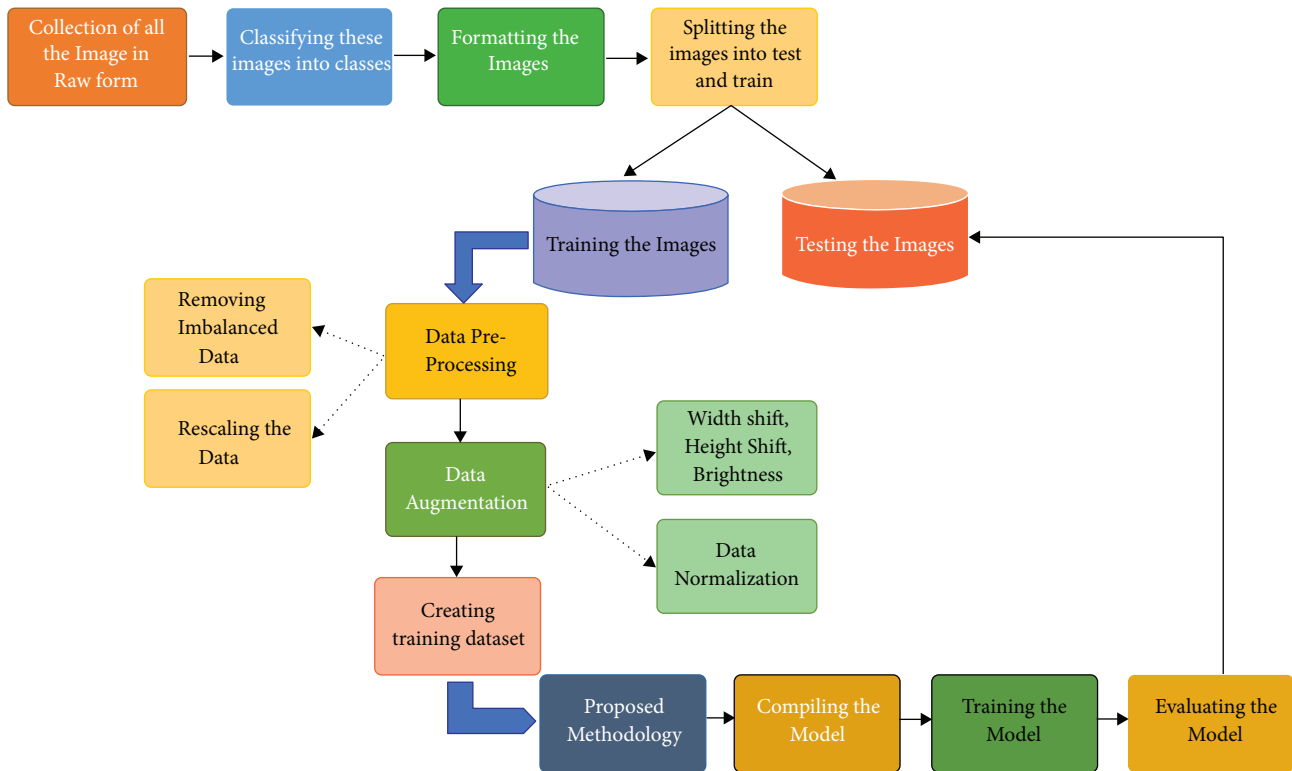


FIGURE 9: The overall flow of the work.

TABLE 1: Statistics of Keras pretrained models.

Model	Size (MB)	Top-1 acc	Top-5 acc	Parameters	Depth
Xception	88	0.790	0.945	22,910,480	126
VGG16	528	0.713	0.901	138,357,544	23
ResNet50	98	0.749	0.921	25,636,712	—
InceptionV3	92	0.779	0.937	23,851,7841	159
MobileNet	16	0.704	0.895	4,253,864	88
EfficientNet	29	0.810	0.922	19,466,823	—

models on the given dataset. So, the custom-training begins in the next section. The training was carried out on a 32 GB NVIDIA Quadro P1000 GPU with a learning rate of 0.001.

3.2.3. Model Compilation. This section includes a detailed analysis of how to perform the complex training process to produce state-of-the-art results. Transfer learning is the only choice to custom train the selected Keras pretrained models. So, what transfer learning is and how to perform it is explained in the below section.

(1) Transfer Learning. Transfer learning refers to the situation when the knowledge learned in one task or domain is reused to improve the generalization in another domain. From machine learning's perspective, it can be defined as reusing the saved weights of any pretrained model to improve the accuracy or to custom-train your model. To use the weights of any pretrained model, for example, VGG16, EfficientNet, some modifications have to be made to make

the model compatible with training on another dataset. The changes performed on the neural network are elaborated in the next part.

The EfficientNet is a convolutional neural network architecture as shown in Figure 10. That uniformly scales all the depth, width, and resolution dimensions. Generally, the model is made wide, deep, and high resolution. This network is scaled up more efficiently, so, gradually, everything is increased. The network consists of 7 blocks, and each one of these blocks further several subblocks. So, these subblocks additionally contain the layers that are the architecture's main building blocks. What modifications are made in this architecture is explained in the next section.

(2) Modifications in Pretrained Models. Three modifications are made to make the model ready to train in the given case. Those modifications are briefly explained below.

(i) Input layer modification

The input layer is changed considering the dimensions and size of the input images. In the current case, the images are of the size (64, 64), and the ImageDataGenerator class receives the input shape parameter to automatically prepare the input layer of the model to initialize the training process.

(ii) Output layer modification

The output layer is modified depending upon the number of classes. The number of neurons is equivalent to the number of classes at the output layer.

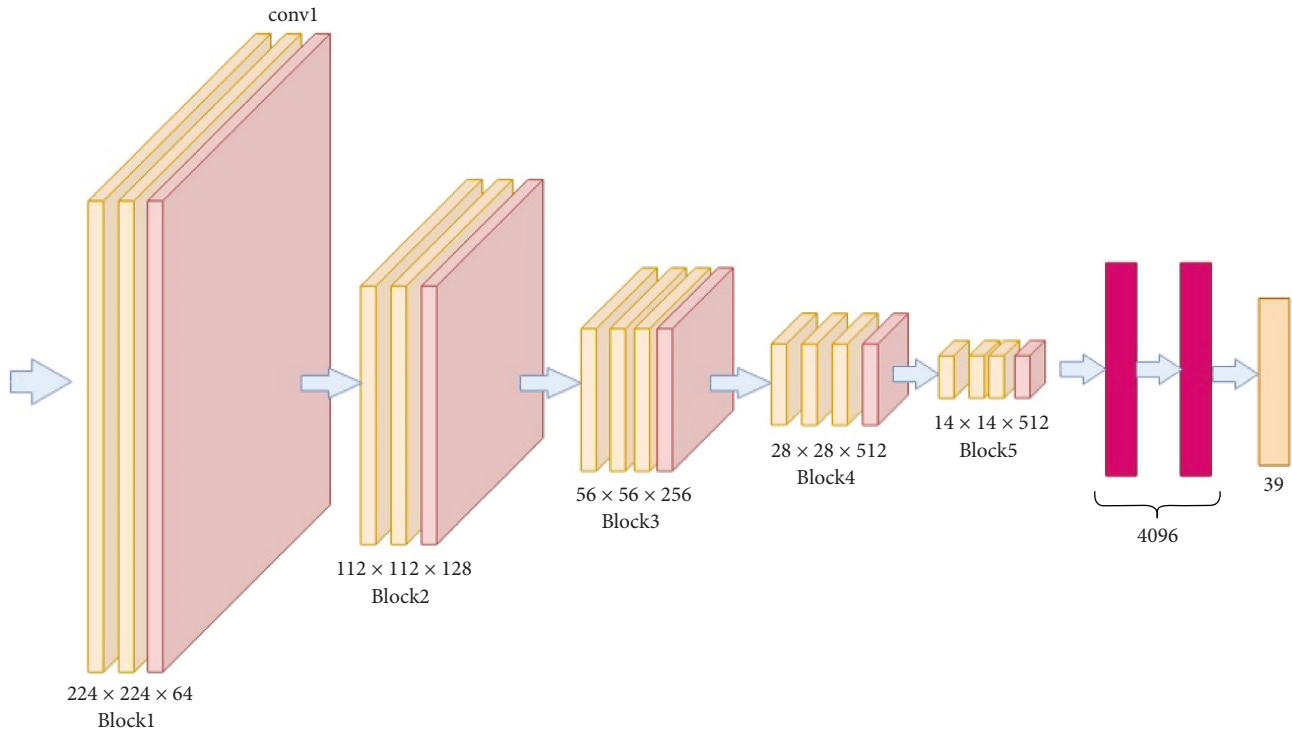


FIGURE 10: EfficientNetB4 architecture.

(iii) Addition of layers

Some dense (fully connected) layers are added at the bottom of these ready-made architectures just before the output layer to make the model more effective and suitable for use following the complexity and the format of the dataset.

(3) *Optimizer*. An optimizer is a final argument required to compile the model before training phenomena. There are different variants of optimizer available in the Keras library like SGD (stochastic gradient descent), RMS (root mean square), Adam, and so on. For hand gesture recognition, Adam [45] is used. The “Adam” optimizer is used to reduce the loss calculated after each epoch while training. This optimizer uses the stochastic gradient descent method that is based on the adaptive estimation of first-order and second-order moments. This method is computationally efficient, occupies less memory, is invariant to diagonal rescaling of gradients, and is best suited for the problems that require complex processing in terms of data/parameters.

(4) *Loss Function*. The loss function is the necessary argument used in the model compilation. The loss function calculates the training or/and validation losses after each epoch during the training phenomena. This measurement provides the level of goodness that shows how well the model is being trained. An increase in loss degrades the model performance, and a decrease in loss optimizes the model performance. There are several built-in classes available in the Keras library for calculating the loss during training. The

selection depends upon the nature of the dataset. In the case of image classification having more than two classes to predict, the “categorical_crossentropy” class is used. This class computes the cross-entropy loss between the ground truth values and the predicted values resulting from model predictions.

(5) *Training Callbacks*. The callbacks are used to perform specific actions at different stages of the training process. These are useful when a developer wants to save model information during or after the training process. These callbacks can be performed before and after the single batch, start or end of an epoch, and so on. The Keras library provides various callbacks, but in this case, few callbacks are considered to be used during the training of the model. These callbacks have their specific functionality and purpose, briefly explained below.

(6) *Model Checkpoint*. This callback is used to save the Keras model or the model weights after some intervals during the training process. The save model file can be used further to load and start the training again or for testing or evaluation purposes. This callback can be used in several ways, providing the optional arguments to the callback class. The options are described precisely below. Whether to save the model possessing the best performance or to save the model file after each epoch, disregarding the model performance. In the specific case, the best model file is protected if the model is improved as compared to past versions. The callback can only be used based on the monitored quantity. The quantity to be monitored and whether it should be maximized or minimized. The monitored amount can have four options:

train_accuracy, train_loss, validation_accuracy, and validation loss. In this case, validation_accuracy is termed as a monitored quantity. The callback also provides the option of at what frequency it should save the model file. The model checkpoint file is saved after each epoch analyzing the validation accuracy. So, there are several other options available to use this callback, but the above-mentioned options are used in the given case.

(7) *Early Stopping*. This callback is used to stop the training process automatically when model performance stops improving up to a specific limit based on some monitored quantity. As previously mentioned, the amount monitored in the given case is validation accuracy. The training process terminates when the validation accuracy stops improving up to a certain number of epochs. Several optional arguments are used to perform this early stopping, and those options are explained. The validation accuracy (monitored quantity) qualifies to be improved when increased with the minimum change. This minimum change can be passed as a parameter to the early stopping class as a min_delta argument. This min_delta option controls the threshold of change to be qualified as improved validation accuracy. The patience option controls when the training process is terminated automatically. The training automatically ends when the model performance starts degrading for the defined number of epochs. This termination is caused when model degradation crosses the patient value specified in the early stopping class.

(8) *CSV Logger*. This callback is used to save the training statistics in a file at runtime during training phenomena. The result of each epoch is held in that file. In addition, a comma-separated log file is used to save the results after each epoch in the given case. So, the callbacks mentioned above are passed as an array to fit() function to apply these callback operations to hunt the most optimal model better and save the evaluation matrices. After defining the training and validation generators to make the dataset ready to train, finalizing the optimizer, loss function, and applying the training callbacks, the model compiles successfully. After successful compilation, the Keras model is now ready to train, which is explained in the next section.

4. Experiments and Discussion

4.1. *Model Training*. At this stage, the model instance is fitted to the fit() function to start the training process. This function trains the model for a fixed number of epochs (iterations on a dataset) using training and validation generators that incorporate the preprocessed images and other required attributes as mentioned in the preceding sections. The model had been trained for 25–30 epochs in almost 10 hours, and each epoch took 2000 steps to complete. The number of steps per epoch depends upon the batch size and the training number of images. For example, the number of training images is 128000, and the batch size is 64. The number of steps per epoch is calculated by dividing the number of training images by batch size, equivalent to 2000 steps in the given case.

4.2. *Performance Metric for Evaluating the Model*. Also, we used other different evaluation metrics as the precision and recall and *F1*-score to evaluate our model concerning each class individually from the 32 Arabic alphabet sign classes as shown in Table 2.

4.2.1. *Precision*. It is also known as the Positive Predictive Value. Accuracy is defined as the proportion of correct predictions divided by the total number of correct class values projected. Equation (1) is used to calculate precision.

$$\text{Precision} = \frac{(\text{True Positive})}{(\text{True Positive} + \text{False Positive})} \quad (1)$$

4.2.2. *Recall*. It is sometimes referred to as vulnerability. Recall is defined as the proportion of correct predictions divided by the number of correct class values. Equation (2) is used to calculate recall.

$$\text{Recall} = \frac{(\text{True Positive})}{(\text{True Positive} + \text{False Negative})} \quad (2)$$

4.2.3. *F1-Score*. The *F*-score or *F*-measure is another name for the *F*-score. The *F1*-score represents the balance between precision and recall. Only when the precision and recall numbers both are good does the *F1*-score grow high. *F1*-score values array from 0 to 1, with the greater the number, the greater the classification accuracy.

F1 – score is calculated by Equation,

$$F1 - \text{score} = \frac{(2 * \text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (3)$$

4.3. *Model Evaluation*. After the model is trained, testing is required to measure the model's real performance on unseen data that the model has not encountered yet. The scikit-learn library provides different programmatic approaches to test the performance of the trained model. The statistical evaluation is done using two methods: confusion matrix and classification report. To describe the performance of the classification model on a test dataset, the classification report is represented in Figure 11.

Figure 11 illustrates the representation of the leading classification matrix on a per-class basis. This visual report gives better and deeper intuition about the classifier's behavior, showing the trained model's functional weaknesses in many analytical aspects. Here, the support column shows the count of test images per class; for example, all classes constitute 1000 test images. The total test samples are 32000. *F1*-score is the mean of precision and recall. The ability of the classifier to find all positive instances (correct predictions) is defined by the recall column numerically. All the classes show the true predictions of more than 95 percent in the recall column except five classes. The report shows the testing accuracy to be 95 percent, which is the real predictive

TABLE 2: Comparative study.

Reference	Method	No. of samples (train/test)	Accuracy
[46]	Deep learning using R-CNN	6300/1570	93
[47]	Semantic segmentation—CNN	43239/10810	88
[48]	Keras pretrained models—CNN	526190/16000	87
Our approach	EfficientNetB4	128000/32000	95

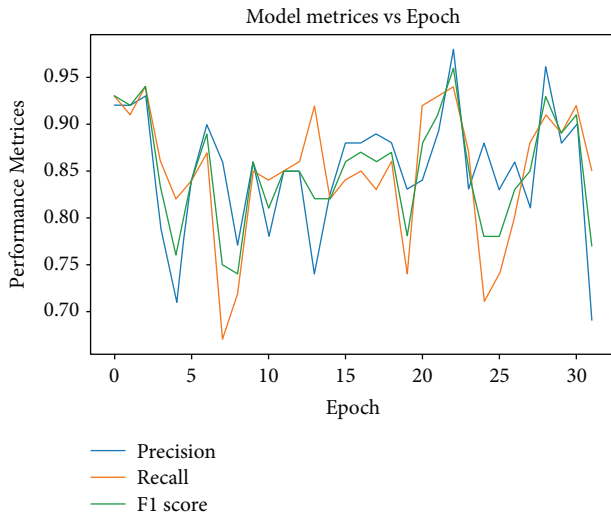


FIGURE 11: Classification report.

result of our classifier. The accuracy of the other results is explained in Table 3.

To find the best-suited model for the given case, several Keras pretrained models are taken into the trial. Eventually, EfficientNetB4 has the best performance in terms of accuracy and loss. To analyze the best-suited model for the given case, it is recommended to plot the graphs for accuracy and loss. So, the graphical approach is used to represent the whole training history of the model with epoch count on the x -axis and the accuracy/loss on the y -axis. The parameters on the y -axis include validation loss, validation accuracy, training loss, and training accuracy. Figures 12 and 13 provide a deep insight into how well the EfficientNetB4 model is trained. It can be seen that the accuracies in Figure 12 and the losses in Figure 13 are converging towards each other steadily up to the 10th epoch. After the 10th epoch, both the accuracies and losses diverge from each other and thus, indicating that the model has learned the weights well. So, the model stops until the 25th epoch to avoid overfitting because the model has known the input features to better classify the unforeseen hand gestures. The behavior can be seen graphically as below.

After having the graphical analysis on the training and the validation performance of the model, the following general step is to test the model on unforeseen data. The random and unexpected evaluation tests the actual intelligence of the trained model about how well the model has learned from the input information. This unpredictable behavior is explained precisely in the next section.

TABLE 3: Evaluations.

Terms	Precision	Recall	F1-score
Accuracy	0.956	0.962	0.95
Macro average	0.95	0.95	0.95
Weighted average	0.95	0.95	0.95

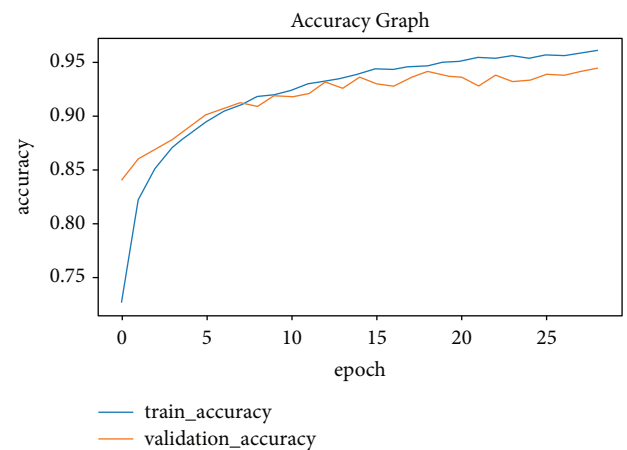


FIGURE 12: Accuracy graph.

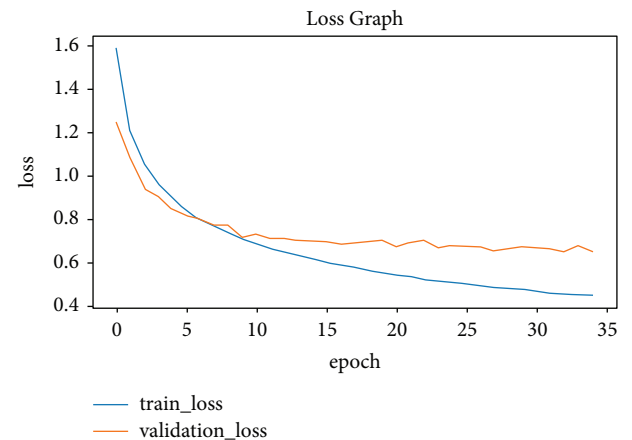


FIGURE 13: Loss graph.

4.4. Comparative Analysis. The previous work done with hand gesture recognition is not so generalized and authentic to use in different environments. Also, based on the dataset, previous papers published include the dataset consisting of not more than 20 classes, but in the given case, the dataset contains around 32 classes constituting nearly 160000 images. In most cases, considering the previous work, the dataset is converted into grayscale images and this dataset

TABLE 4: Comparative analysis (current work and previous work).

Current work	Past work
More generalized solution	Specialized solution [49]
31 classes	Less than 20 classes [49]
The model trained on RGB images	The model trained on grayscale images [49]
Applied data augmentation techniques	No data augmentation techniques [50]
It can be applied in real time	In most cases, not applicable in real time [51]

transformation sometimes degrades the model performance. Secondly, several data augmentation techniques are applied in the given case to make the solution adaptable to different environments. Concerning generalization versus specialization, the model trained in the current issue is more generalized than the models analyzed in past papers. The datasets and approaches explored in the previous work indicate that the solution is not adaptable and generalized. Therefore, the specialization problem is now resolved because the model is trained in a diverse environment. For the applicability considering the given case, the solution is more applicable than the past work. It can be adapted in real-time environments as well. Briefly, the comparative analysis is summarized in Table 4.

As shown in the above Table 4, the proposed method is better than the 2 methods and better than many in terms of classes.

The novelty of the proposed work is listed as follows: The proposed work is hand gesture recognition using the simple and efficient classifier. It includes the process of retraining the pretrained TensorFlow architectures. It includes the absence of the sensor hardware. Most of the approaches used to perform Sign Language Recognition includes wearing hand data gloves for the acquisition of hand gesture data. Instead, the current approach does not include any hardware but the mandatory camera. The proposed method is highly efficient and feasible for real-time applications. This method includes the current system, which becomes more applicable when hand detection is added. Adding hand detection and tracking stage makes this application more adaptable and comprehensive. The proposed system has a low latency of classifying the hand gestures.

5. Discussion

The whole workflow is discussed following the steps of data preprocessing and model training and evaluations. Data augmentation has played a vital role in preventing the training models from being overfitted, thus improving the models' overall performance on the unforeseen dataset. For data augmentation on the image's dataset, 5000 augmented images are generated using 1000 original images per class. Considering the number of classes that is 32, 5000 * 32 images are generated using 1000 * 32 original dataset. Consequently, 160000 augmented images are generated using 32000 original normalized images and, thus, played a vital role in the prevention of degradation of the models.

Other than this, the preprocessing techniques and architecture modifications applied are the crucial steps in successful sign language recognition. The results accomplished are marginally better than the past works. Considering the current and the past work as summarized in Table 4, this solution is more generalized and better as it performs even better on the test (unforeseen) dataset. But most of the past papers include the accuracies for the specific case without augmenting the dataset. Secondly, the model trained in the current work is based on images containing three colors' spaces instead of grayscale images to make the training process faster.

On the contrary, most of the past papers include training on grayscale images. Focusing on the problem domain, the grayscale images may not perform better considering the high similarity index between the given dataset classes. Thirdly, the current work can be made applicable in real time if hand detection and tracking are made possible in this case. So, these are the few reasons why the current work is better than most of the work done in the past on Arabic sign language recognition or classification problems or hand gesture recognition simply. Knowingly, several pretrained models are trained in the flow, but the best fit is concluded in the final and next section.

6. Conclusions

The study is concluded to be the best in its way for Arabic sign language recognition. Considering the steps of the workflow, the very first step is image acquisition. The image acquisition involved the acquisition of the original image taken from the test(unforeseen) dataset to be preprocessed further in the next step. The preprocessing step involves several substeps. The first substep in the preprocessing part is to rescale the given original image that matches the input shape of the model architecture. The input shape was finalized to be $(64 \times 64 \times 3)$, where 3 indicates the number of color spaces or the dimensions of the input image. After rescaling the image, it is normalized by limiting the pixel values between the range (0, 1). The normalization is performed on the input image so that model may find it easy to extract features and propagate them in between the layers of the architecture for fast prediction with a low latency rate. After normalization, the model receives the preprocessed input image and tries to predict the meaningful pattern on which the model is trained at. Finally, the model attempts to predict the relevant trained patterns of the input image through classification. The classification is made by considering the best-known hand gesture with the highest probability of it happening at most. That class with the highest change is the final prediction made by the trained model. So far, several pretrained models have been tried on various given datasets. Most of them performed pretty usually, and the EfficientNetB4 is considered the best fit for the case in the final stage. Considering the complexity of the dataset, models other than EfficientNetB4 do not perform well due to their lightweight architecture. EfficientNetB4 is a heavy-weight architecture that possesses more complexities comparatively. However, due to its ability to perform on the

extensive dataset consisting of a high number of classes, it is the high performer among other pretrained models. The best model is exposed with a training accuracy of 98 percent and a testing accuracy of 95 percent. These accuracies are represented in Figures 12 and 13 as the model is evaluated at the final stage. As a future recommendation, we would like to combine various techniques to handle single-hand gesture recognition. The multiple techniques could be MobileNet and ResNet50 architectures. Another recommendation would be to apply these techniques to detect the gestures from both hands.

Data Availability

The dataset used in this study is taken from [43].

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors extend their appreciation to the Researchers Supporting Project no. RSP-2021/322, King Saud University, Riyadh, Saudi Arabia.

References

- [1] P. Shukla, A. Garg, K. Sharma, and A. Mittal, "A DTW and fourier descriptor based approach for Indian sign language recognition," in *Proceedings of the 2015 Third International Conference on Image Information Processing*, pp. 113–118, ICIIIP), Wagnaghat, India, December 2015.
- [2] R. Kushalnagar, "Deafness and hearing loss," in *Web Accessibility*, pp. 35–47, Springer, Berlin, Germany, 2019.
- [3] C. R. de Souza and E. B. Pizzolato, "Sign language recognition with support vector machines and hidden conditional random fields: going from fingerspelling to natural articulated words," in *Proceedings of the International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 84–98, New York, NY, USA, July 2013.
- [4] S. S. Kumar, T. Wangyal, V. Saboo, and R. Srinath, "Time series neural networks for real time sign language translation," in *Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications*, pp. 243–248, ICMLA), Orlando, FL, USA, December 2018.
- [5] A. S. Nikam and A. G. Ambekar, "Sign language recognition using image based hand gesture recognition techniques," in *Proceedings of the 2016 Online International Conference on green Engineering and Technologies*, pp. 1–5, IC-GET), Coimbatore, India, November 2016.
- [6] M. A. Rahaman, M. Jasim, M. H. Ali, and M. Hasanuzzaman, "Real-time computer vision-based Bengali sign language recognition," in *Proceedings of the 2014 17th International Conference on Computer and Information Technology (ICCIT)*, pp. 192–197, Dhaka, Bangladesh, December 2014.
- [7] B. ChandraKarmokar, K. M. Rokibul Alam, and M. Kibria Siddiquee, "Bangladeshi sign language recognition employing neural network ensemble," *International Journal of Computer Application*, vol. 58, no. 16, pp. 43–46, 2012.
- [8] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign language recognition using convolutional neural networks," in *Proceedings of the European Conference on Computer Vision*, pp. 572–578, Zurich, Switzerland, September 2014.
- [9] T. F. Ayshee, S. A. Raka, Q. R. Hasib, M. Hossain, and R. M. Rahman, "Fuzzy rule-based hand gesture recognition for Bengali characters," in *Proceedings of the 2014 IEEE International Advance Computing Conference (IACC)*, pp. 484–489, Gurgaon, India, February 2014.
- [10] M. Mustafa, "A study on Arabic sign language recognition for differently abled using advanced machine learning classifiers," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 4101–4115, 2021.
- [11] M. Maraqa and R. Abu-Zaiter, "Recognition of Arabic Sign Language (ArSL) using recurrent neural networks," in *Proceedings of the 2008 First International Conference on the Applications of Digital Information and Web Technologies (ICADIWT)*, pp. 478–481, Ostrava, Czech Republic, August 2008.
- [12] R. Alzohairi, R. Alghonaim, W. Alshehri, S. Aloqeely, M. Alzaidan, and O. Bchir, "Image based Arabic sign language recognition system," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 3, 2018.
- [13] A.-G. A.-R. Abdel-Samie, F. A. Elmisery, A. M.Brisha, and A. H. Khalil, "Arabic sign language recognition using kinect sensor," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 15, no. 2, pp. 57–67, 2018.
- [14] Z. Hu, Y. Zhang, Y. Xing, Y. Zhao, D. Cao, and C. Lv, "Toward human-centered automated driving: a novel spatiotemporal vision transformer-enabled head tracker," *IEEE Vehicular Technology Magazine*, pp. 2–9, 2022.
- [15] Z. Hu, Y. Zhang, Y. Xing, Y. Zhao, D. Cao, and C. Lv, "Towards Human-Centered Automated Driving: A Novel Spatial-Temporal Vision Transformer-Enabled Head Tracker," *IEEE Vehicular Technology Magazine*, vol. 1, 2022.
- [16] A. A. Youssif, A. E. Aboutabl, and H. H. Ali, "Arabic sign language (arsl) recognition system using hmm," *International Journal of Advanced Computer Science and Applications*, vol. 2, no. 11, 2011.
- [17] M. Abdo, A. Hamdy, S. Salem, and E. M. Saad, "Arabic alphabet and numbers sign language recognition," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 11, pp. 209–214, 2015.
- [18] N. El-Bendary, H. M. Zawbaa, M. S. Daoud, A. E. Hassanien, and K. Nakamatsu, "Arslat: Arabic sign language alphabets translator," in *Proceedings of the 2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, pp. 590–595, Krakow, Poland, October 2010.
- [19] M. A. Almasre and H. Al-Nuaim, "A real-time letter recognition model for Arabic sign language using kinect and leap motion controller v2," *Int. J. Adv. Eng. Manag. Sci.* vol. 2, no. 5, Article ID 239469, 2016.
- [20] M. ElBadawy, A. S. Elons, H. A. Shedeed, and M. F. Tolba, "Arabic sign language recognition with 3d convolutional neural networks," in *Proceedings of the 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 66–71, Cairo, Egypt, December 2017.
- [21] S. Hayani, M. Benaddy, O. El Meslouhi, and M. Kardouchi, "Arab sign language recognition with convolutional neural networks," in *Proceedings of the 2019 International Conference of Computer Science and Renewable Energies (ICCSRE)*, pp. 1–4, Agadir, Morocco, July 2019.

- [22] B. Kayalibay, G. Jensen, and P. van der Smagt, "CNN-based segmentation of medical imaging data," 2017, <https://arxiv.org/abs/1701.03056>.
- [23] M. S. Hossain and G. Muhammad, "Emotion recognition using secure edge and cloud computing," *Information Sciences*, vol. 504, pp. 589–601, 2019.
- [24] M. M. Kamruzzaman, "E-crime management system for future smart city," in *Data Processing Techniques and Applications for Cyber-Physical Systems (DPTA 2019)*, pp. 261–271, Springer, Berlin, Germany, 2020.
- [25] O. K. Oyedotun and A. Khashman, "Deep learning in vision-based static hand gesture recognition," *Neural Computing & Applications*, vol. 28, no. 12, pp. 3941–3951, 2017.
- [26] S. Ahmed, M. Rafiqul Islam, J. Hassan et al., "Hand sign to Bangla speech: a deep learning in vision based system for recognizing hand sign digits and generating Bangla speech," 2019, <https://arxiv.org/abs/1901.05613>.
- [27] Z. Hu, Y. Hu, J. Liu, B. Wu, D. Han, and T. Kurfess, "A CRNN module for hand pose estimation," *Neurocomputing*, vol. 333, pp. 157–168, 2019.
- [28] U. Côté-Allard, C. L. Fall, A. Drouin et al., "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 760–771, 2019.
- [29] Z. Hu, Y. Xing, C. Lv, P. Hang, and J. Liu, "Deep convolutional neural network-based Bernoulli heatmap for head pose estimation," *Neurocomputing*, vol. 436, pp. 198–209, 2021.
- [30] Z. Hu, Y. Hu, J. Liu, B. Wu, D. Han, and T. Kurfess, "3D separable convolutional neural network for dynamic hand gesture recognition," *Neurocomputing*, vol. 318, pp. 151–161, 2018.
- [31] Y. Si, S. Chen, M. Li, S. Li, Y. Pei, and X. Guo, "Flexible strain sensors for wearable hand gesture recognition: from devices to systems," *Adv. Intell. Syst.* vol. 4, Article ID 2100046, 2022.
- [32] H. Wang, Y. Zhang, C. Liu, and H. Liu, "sEMG based hand gesture recognition with deformable convolutional network," *Int. J. Mach. Learn. Cybern.*, vol. 1, pp. 1–10, 2022.
- [33] M. M. Alam, M. T. Islam, and S. M. M. Rahman, "Unified learning approach for egocentric hand gesture recognition and fingertip detection," *Pattern Recognition*, vol. 121, Article ID 108200, 2022.
- [34] Y. Chenyi, H. Yuqing, Z. Junyuan, and L. Guorong, "Light-weight neural network hand gesture recognition method for embedded platforms," *High Power Laser and Particle Beams*, vol. 34, no. 3, pp. 1–9, 2022.
- [35] S. Joudaki and A. Rehman, "Dynamic hand gesture recognition of sign language using geometric features learning," *International Journal of Computational Vision and Robotics*, vol. 12, no. 1, pp. 1–16, 2022.
- [36] N. Tubaiz, T. Shanableh, and K. Assaleh, "Glove-based continuous Arabic sign language recognition in user-dependent mode," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 4, pp. 526–533, 2015.
- [37] S. M. S. Al-Buraiky, *Arabic Sign Language Recognition Using an Instrumented Glove*, King Fahd University of Petroleum and Minerals (Saudi Arabia), Saudi Arabia, 2004.
- [38] Z. Hu, Y. Hu, B. Wu, J. Liu, D. Han, and T. Kurfess, "Hand pose estimation with multi-scale network," *Applied Intelligence*, vol. 48, no. 8, pp. 2501–2515, 2018.
- [39] S. M. Halawani, "Arabic sign language translation system on mobile devices," *IJCSNS Int. J. Comput. Sci. Netw. Secur.* vol. 8, no. 1, pp. 251–256, 2008.
- [40] M. Mohandes, M. Deriche, and J. Liu, "Image-based and sensor-based approaches to Arabic sign language recognition," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 4, pp. 551–557, 2014.
- [41] M. Almasre and H. Al-Nuaim, "Comparison of four SVM classifiers used with depth sensors to recognize Arabic sign language words," *Computers*, vol. 6, no. 2, p. 20, 2017.
- [42] Z. Hu, C. Lv, P. Hang, C. Huang, and Y. Xing, "Data-driven estimation of driver attention using calibration-free eye gaze and scene features," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 2, pp. 1800–1808, 2022.
- [43] G. Latif, N. Mohammad, J. Alghazo, R. AlKhalaf, and R. AlKhalaf, "ArASL: Arabic alphabets sign language dataset," *Data in Brief*, vol. 23, p. 103777, 2019.
- [44] Keras, "Deep Learning for humans," 2021, <https://github.com/fchollet/keras>.
- [45] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2014, <https://arxiv.org/abs/1412.6980>.
- [46] R. A. Alawwad, O. Bchir, and M. M. Ben Ismail, "Arabic Sign Language Recognition Using Faster R-CNN," *International Journal of Advanced Computer Science and Applications*, vol. 12, 2021.
- [47] A. Althagafi, G. A. T. Alsubait, and T. Alqurash, "ASLR: Arabic sign language recognition using convolutional neural networks," *IJCSNS International Journal of Computer Science and Network Security*, vol. 20, no. 7, 2020.
- [48] Y. S. Tan, K. M. Lim, and C. P. Lee, "Hand gesture recognition via enhanced densely connected convolutional neural network," *Expert Systems with Applications*, vol. 175, p. 114797, 2021.
- [49] Y. Saleh and G. Issa, "Arabic Sign Language Recognition through Deep Neural Networks fine-tuning," *IJOE*, vol. 16, 2020.
- [50] A. A. Alani and G. Cosma, "ArSL-Cnn: A Convolutional Neural Network for Arabic Sign Language Gesture Recognition," *Indonesian journal of electrical engineering and computer science*, vol. 22, 2021.
- [51] E. K. Elsayed and D. R. Fathy, "Sign language semantic translation system using ontology and deep learning," *International Journal of Advanced Computer Science and Applications*, vol. 11, 2020.

Research Article

Dynamic and Static Features-Aware Recommendation with Graph Neural Networks

Ninghua Sun ^{1,2} Tao Chen ¹ Longya Ran ¹ and Wenshan Guo ¹

¹School of Public Administration, Huazhong University of Science and Technology, Wuhan 430074, China

²Innovation Institute, Huazhong University of Science and Technology, Wuhan 430074, China

Correspondence should be addressed to Tao Chen; chentao15@163.com

Received 16 March 2022; Revised 27 March 2022; Accepted 28 March 2022; Published 21 April 2022

Academic Editor: Zhongxu Hu

Copyright © 2022 Ninghua Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recommender systems are designed to deal with structured and unstructured information and help the user effectively retrieve needed information from the vast number of web pages. Dynamic information of users has been proven useful for learning representations in the recommender system. In this paper, we construct a series of dynamic subgraphs that include the user and item interaction pairs and the temporal information. Then, the dynamic features and the long- and short-term information of users are integrated into the static recommendation model. The proposed model is called dynamic and static features-aware graph recommendation, which can model unstructured graph information and structured tabular data. Particularly, two elaborately designed modules are available: dynamic preference learning and dynamic sequence learning modules. The former uses all user-item interactions and the last dynamic subgraph to model the dynamic interaction preference of the user. The latter captures the dynamic features of users and items by tracking the preference changes of users over time. Extensive experiments on two publicly available datasets show that the proposed model outperforms several compelling state-of-the-art baselines.

1. Introduction

The amount of information on the Internet continues to grow rapidly, and determining useful information has become increasingly difficult. Fortunately, the advancement of recommender systems can substantially help people deal with the information overload problem. Collaborative filtering (CF) is one of the most famous methods in recommendation algorithms. Therefore, collaborative learning latent representations of users and items from user-item interactions is an important step in CF-based models. However, poor latent representations of users and items remain the factors limiting further performance.

Therefore, researchers have adopted different methods to capture latent representations. Till now, the most commonly used approach for CF is to learn latent features in the embeddings space generated from the user-item rating matrix, such as matrix factorization [1] and deep learning-based CF

[2–4]. Some researchers [5, 6] use a bipartite graph to represent user-item interactions to further enhance the latent representations; hence, the topological features of the graph are introduced through graph neural networks (GNNs) [7]. The underlying assumption in leveraging the bipartite graph as input to obtain effective recommendations is as follows: nodes that are connected can spread information by aggregating their neighbors, thereby potentially contributing to capturing high-order features.

Latent representations obtained from dynamic user-item interactions serve as another method. Traditional CF usually defines a decay function of temporal information [8, 9], such as the exponential decay function $e^{-\omega t}$, to capture these dynamic features, while graph-based CF obtains a series of user-item bipartite graphs based on interaction time [10]. The underlying assumption in using temporal information is that the behaviors of users on items are a dynamic interactive process; consequently, the long- and short-term preferences of users are captured.

It is unclear, however, which of these approaches—static recommendation versus dynamic recommendation—is better for predicting user preference on items. The former ignores that user preference is dynamic, thus changing over time. The latter usually requires more parameters and training time than the static recommendation, which limits its application. Furthermore, the introduction of temporal information may bring additional noise, which can hinder the performance and scalability of the model. Two important problems must be solved to deal with these challenges:

- (1) How to represent the behaviors of users with a dynamic graph? Temporal information is vital for capturing the dynamic preference of users. To avoid introducing additional noise data, utilizing temporal information data more efficiently should be priorities.
- (2) How to obtain the dynamic features simply and swiftly? In addition to the interaction pairs of users and items, the dynamic graph also includes side information (e.g., temporal information). However, additional information will introduce an increase in the number of parameters and high computational complexity.

To this end, a simple and effective graph-based algorithm is proposed to introduce dynamic features into a static recommender system called dynamic and static features-aware graph recommendation (DSAGR). Firstly, rather than simply timestamps, the dynamic graph of users and items is constructed based on Takens' time embedding theorem [11] to use temporal information efficiently. This work employs the graph convolution network (GCN) [7] to learn the long- and short-term preferences of users because of the expressive graph-based models. Then, a novel module is proposed, that is, the dynamic sequence learning module, to transform the unstructured dynamic graph to structured sequence data to decrease the dynamic model complexity. In particular, convolutional neural networks (CNNs) [12, 13] are used to capture the dynamic features from the sequence data. Finally, the dynamic features and dynamic preference are integrated to obtain the predictor for each user. Our main contributions are as follows.

- (1) This work can simply and swiftly capture the dynamic features from the constructed dynamic graph.
- (2) A novel hybrid model is proposed in this work, which can easily capture the users' dynamic preferences.
- (3) An offline experiment is performed on real-world datasets. The results show that the proposed model successfully performs the personalized recommendation task.

The rest of the paper is organized as follows. Section 2 elaborates on relevant research. Section 3 presents the proposed method, while Section 4 discusses the empirical study on the public datasets. Finally, Section 5 contains the conclusions.

2. Related Work

Collaborative filtering- (CF-) based recommendation aims to predict the preference of users and then return top- N items of the user interests. Heuristic works, such as item- and user-based models, predict the preferences of users on items based on the k -nearest neighbor algorithm [14]. Model-based approaches usually learn the user and item with low-rank latent representations through matrix factorization [1]. The inner product between the two lower-rank vectors is then used to obtain the probability of the user clicking on the item.

Furthermore, deep learning has been shown to be particularly well suited to representation learning tasks [15, 16]. Therefore, many deep learning techniques have recently allowed CF to have expressive representation vectors from the historical behaviors of users, such as the multilayer perceptron (MLP), autoencoder (AE), recurrent neural network (RNN), and CNN. Many researchers often consider the combinations of matrix factorization and deep learning techniques for CF recommendation. MLP-based model Wide&Deep [17] captures linear and nonlinear latent features effectively. NeuMF [2] integrates MLP and MF to model high- and low-order interaction features. Furthermore, AE is used for recommendation tasks. Work [18] employs a denoising recurrent AE network and then generalizes it to the CF setting. RNN has been widely used for recommendation due to its excellent performance in modeling sequential data. The variants of RNN, such as long short-term memory (LSTM) [19] and gated recurrent unit (GRU) networks [20], are often employed in practice to overcome the vanishing gradient problem. For instance, work [21] uses LSTM to model the long- and short-term preferences of users. CNN is also a powerful tool [15]. Work [22] uses two parallel CNNs to learn deep representations of users and items. Work [23] integrates CNN and GRU networks to obtain distributed representations of users and items. These representations are then used to regularize the generation of latent features in matrix factorization.

In the last few years, GNN has been widely recognized as a state-of-the-art approach because of its successful applications in recommendation tasks [24]. GNN can effectively learn the structural representations of nodes by aggregating their neighborhood information. A pooling operation is typically used to output the node embeddings after an aggregation function. Many graph-based models are also proposed by using different aggregation and pooling functions such as GCN [25], GraphSAGE [26], and Graph Isomorphism Network (GIN) [27]. Among these models, the most popular recommendation method is LightGCN [6] coupled with NGCF [5]. LightGCN is an effective simplified version of the NGCF by omitting the transformation mechanism and applying the sum-based pooling layer. Some researchers also consider dynamic representation learning to model data. Work [28] employs matrix perturbation to model the changes in graphs, such as the adjacency matrix. Work [29] constructs the user-item interaction graph dynamically based on the users and items embeddings to improve the diversity of recommendations. Furthermore,

many graph-based algorithms [24] have been proposed to enrich the presentation of users and items with other auxiliary information [30, 31]. Therefore, this work tries to introduce dynamic features as side information to improve the recommendation performance.

These graph-based models have verified their superiority for the recommendation task. However, these models mainly focus on constructing static graph-based recommendation models without considering their combinations with dynamic graph features. As far as we know, there is no study to introduce dynamic graph features into a graph-based recommendation framework.

3. Proposed Method

In this part, the proposed DSAGR method is presented, and its framework is illustrated in Figure 1. Four components are included in the framework: (1) dynamic graph construction aims to convert the behaviors of users into a dynamic graph; (2) dynamic preference learning module is to learn the long- and short-term preferences of users; (3) dynamic sequence learning module aims to capture the user and item sequence features as side information; and (4) prediction layer is to obtain the predictors.

3.1. Dynamic Graph Construction. Given a user set U , an item set I , and a set of time stamps $T = \{t_1, t_2, t_3, \dots\}$, the graph of the user-item interaction at the time stamp t_1 can be defined as $G_{t_1} = (U \cup I, \mathcal{E}_{t_1})$, where $U \cup I$ is the set of nodes, and edge $e \in \mathcal{E}_{t_1}$ represents the interaction between the user and the item at the time $t_1 \in T$. Therefore, the interactions of users and items can be seen as a time series, that is, $\{G_{t_1}, G_{t_2}, G_{t_3}, \dots\}$. Figure 2(b) shows different graphs at five different time stamps.

To understand the behaviors of users with the effects of temporal information, several time slices of user-item interactions are generated based on Takens' time embedding theorem [11] using a given delay factor. Considering the following example: given five timestamps [1–5], we assume that the delay factor is equal to 1 and the number of the time slices is 4. Takens' time embedding theorem indicates that the time series is embedded into R^2 vector space as follows: $[[1, 2], [2, 3], [3, 4], [4, 5]]$. Similarly, the user-item interaction time stamps can also be embedded into the vector space and further be divided into l time slices $[T_1, T_2, T_3, \dots, T_l]$. Therefore, an interaction graph for each time slice can be obtained as previously mentioned. More formally, the obtained interaction subgraphs are denoted as $\{G_{T_1}, G_{T_2}, G_{T_3}, \dots, G_{T_l}\}$.

Figure 2 demonstrates the specific processes. Figure 2(a) presents the example dataset, which is ranked based on the interaction time in the order from small to large. For the sake of convenience, the interaction time is indicated by numbers 1–5. In (b), the user-item interaction graph at different timestamps can be observed. These user nodes are marked dark red, and item nodes are marked lilac color. In (c), Takens' embedding of temporal information generates three

time slices marked by orange color (i.e., T_1 : [1–3], T_2 : [2–4], and T_3 : [3–5]), in which the element is the time ID. The interacted pairs in each time slice constitute a user-item interaction subgraph.

3.2. Dynamic Preference Learning Module. The upper part in Figure 1 shows the dynamic preference learning module. In the recommendation task, the long-term preference of users reflects their inherent features and general preference, which can be learned from all interacted items of users. The short-term signals of the user reflect his/her latest preference. Furthermore, many studies [32] use the latest interaction item embedding and the latest timestamp as short-term information but ignore the dependence on historical interactions. The long- and short-term collaboration can be captured effectively by considering the same layer structure with Siamese and information sharing components [33] on all interaction graph G and the last subgraph G_{T_l} . Siamese networks can naturally introduce inductive biases for invariance modeling because of identical weight-sharing subnetworks. Then, the two graphs can be parameterized using a GNN layer, such as LightGCN [6]. To offer a holistic view of the long-term and short-term collaborative nodes embeddings, we provide the matrix form.

Long-term:

$$\begin{aligned} \begin{pmatrix} L_u \\ L_i \end{pmatrix} &= (D^{-1/2} A D^{1/2}) \begin{pmatrix} E_u \\ E_i \end{pmatrix}, \\ \begin{pmatrix} E_{\text{long},u} \\ E_{\text{long},i} \end{pmatrix} &= \lambda_0 \begin{pmatrix} E_u \\ E_i \end{pmatrix} + \lambda_1 \begin{pmatrix} L_u \\ L_i \end{pmatrix}, \\ A &= \begin{pmatrix} M & 0 \\ 0 & M^T \end{pmatrix}, \end{aligned} \quad (1)$$

where $M \in R^{|U| \times |I|}$ is the user-item rating matrix, in which each element $M_{u,i}$ is 1 if the user u interacted with the item i ; otherwise, it is 0. Then, A is the adjacency matrix of the graph G ; D is a $(|U| + |I|) \times (|U| + |I|)$ diagonal matrix, in which each entry D_{jj} is the number of nonzero entries in the j th row vector of the adjacency matrix A ; $E_u \in R^{|U| \times d}$, $E_i \in R^{|I| \times d}$ are the initial weight matrix of users and items, respectively. $\begin{pmatrix} E_u \\ E_i \end{pmatrix}$ denotes the concatenation of E_u and E_i . $\lambda_0, \lambda_1 \in [0, 1]$, are the defined hyperparameters; $E_{\text{long},u}$ and $E_{\text{long},i}$ are the final representations of users and items for learning long-term preferences.

Short-term:

$$\begin{aligned} \begin{pmatrix} S_u \\ S_i \end{pmatrix} &= (D_{\text{last}}^{-1/2} A_{\text{last}} D_{\text{last}}^{1/2}) \begin{pmatrix} E_u \\ E_i \end{pmatrix}, \\ \begin{pmatrix} E_{\text{short},u} \\ E_{\text{short},i} \end{pmatrix} &= \lambda_0 \begin{pmatrix} E_u \\ E_i \end{pmatrix} + \lambda_1 \begin{pmatrix} S_u \\ S_i \end{pmatrix}, \end{aligned} \quad (2)$$

where A_{last} is the adjacency matrix of the latest subgraph G_{T_l} ; D_{last} is also the diagonal matrix calculated based on A_{last} . $E_{\text{short},u}$ and $E_{\text{short},i}$, respectively, denote the final representation of users and items in the short-term preference learning.

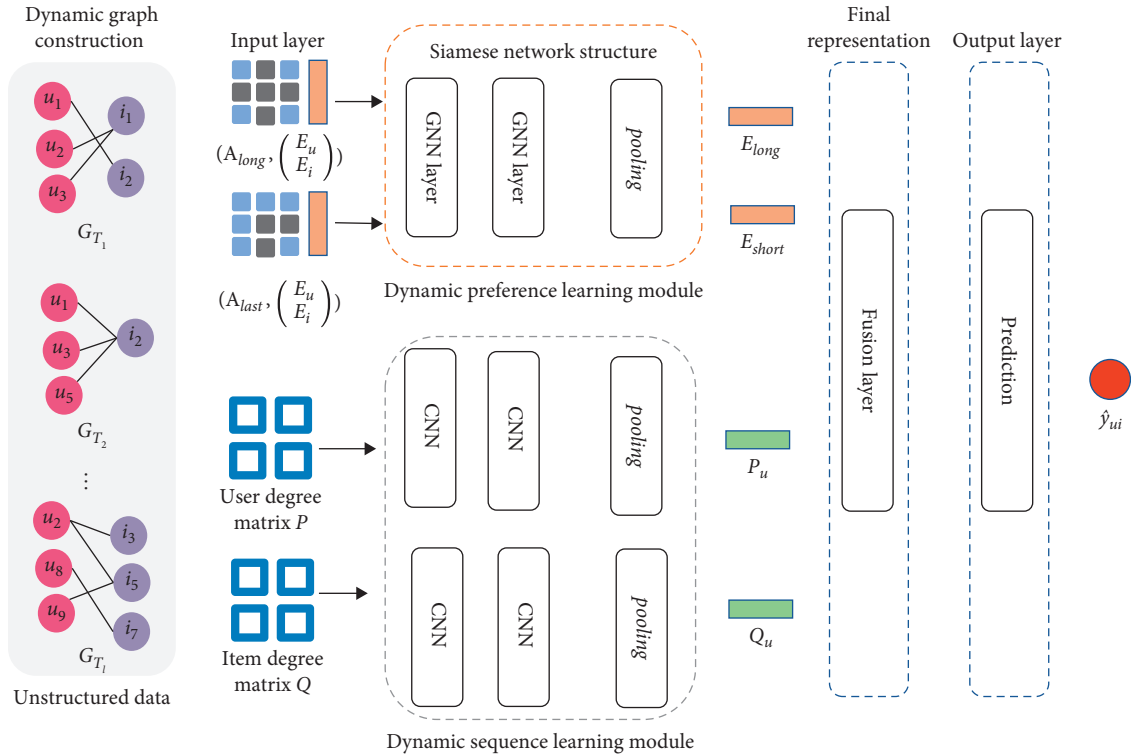


FIGURE 1: Framework of our model.

Interaction time ID	1	2	3	4	5
User ID	1	1	2	1	3
Item ID	1	2	1	3	1

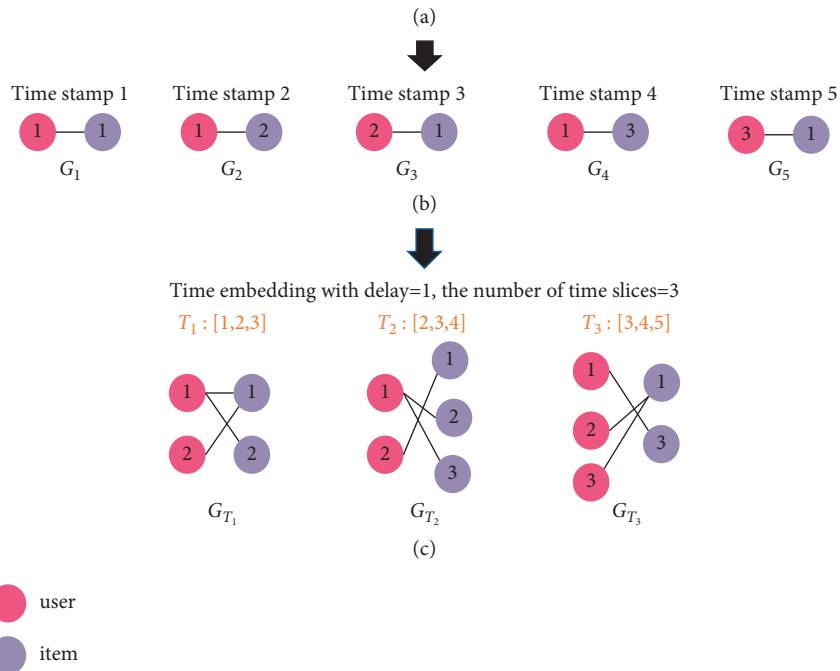


FIGURE 2: The construction of the dynamic graph. (a) Dataset. (b) The graph of user-item interaction at different time stamps. (c) Dynamic subgraphs.

3.3. Dynamic Sequence Learning Module. The degree of the graph is shown to be effective for evaluating the popularity of nodes [34, 35]. Therefore, the degree matrixes of users and items are proposed to track the dynamic changes in the user and item nodes in constructed dynamic graph $\{G_{T_1}, G_{T_2}, G_{T_3}, \dots, G_{T_l}\}$, respectively. For instance, the degree matrix of items is denoted as $Q = \{q_1, q_2, \dots, q_l\}$, in which element q_k is a $|I|$ dimensional vector, its i th element $q_{i,k}$ is the number of edges incident to the item i in the k th subgraph G_{T_k} . And the element $q_{i,k}$ means the popularity of item i in the time slice T_k . Similarly, the user degree matrix is denoted as $P = \{p_1, p_2, \dots, p_l\}$, where $p_l \in R^{|U|}$. Therefore, this study offers a novel means of processing the unstructured graph data and hence may shed light on the task of graph-based recommendation.

The lower part in Figure 1 shows the dynamic sequence features learning modeled by two parallel CNN layers. The input of the module is the obtained user degree matrix and item degree matrix $P \in R^{|U| \times l}, Q \in R^{|I| \times l}$. The CNNs generally comprise a set of convolutional and pooling layers in their architectures. In this work, two 1D-convolutional layers and one pooling layer are designed to learn dynamic features. The first and second convolutional layers with a set of $\{f_1, f_2\}$ filters with the kernel size of τ , shared weights $w_1 \in \mathbb{R}^{f_1 \times 1 \times \tau}, w_2 \in \mathbb{R}^{f_2 \times f_1 \times \tau}$, as shown in the following equations.

$$g_t = p_t * w_1, \quad (3)$$

$$h_t = g_t * w_2, g_t \in R^{|U| \times f_1}, \quad (4)$$

where $p_t \in R^{|U|}$ is the column vector of user degree matrix $P \in R^{|U| \times l}$, $*$ is the convolution operator, and $h_t \in R^{|U| \times f_2}$ denotes a feature matrix for all users. After the 1D-convolutional operation, l feature matrixes can be obtained. Inspired by graph-based models [5, 6], the weighted sum operator is designed as the pooling layer and then normalized by the sigmoid function σ . The output is shown in the following equation.

$$P_u = \sigma(h_1 + h_2 + \dots + h_l). \quad (5)$$

Analogously, the items' degree representations are defined as follows.

$$\begin{aligned} g'_t &= q_t * w'_1, t = 1, 2, \dots, l, \\ h'_t &= g'_t * w'_2, g'_t \in R^{|I| \times f_1}, t = 1, 2, \dots, l, \\ Q_i &= \sigma\left(\sum_{t=0}^l h'_t\right), h'_t \in R^{|I| \times f_2}, \end{aligned} \quad (6)$$

where $q_t \in R^{|I|}$ is the column vector in item degree matrix Q , $*$ is the convolution operator, and $w'_1 \in \mathbb{R}^{f_1 \times 1 \times \tau}, w'_2 \in \mathbb{R}^{f_2 \times f_1 \times \tau}$ are shared factors.

3.4. Prediction Layer. The embeddings and degree representations of nodes of the user and item are obtained after the dynamic preference learning module and dynamic

sequence learning module. Then, a fusion layer is defined to learn the final representations:

$$E_u^* = (E_{\text{long},u} + E_{\text{short},u}, P_u), E_i^* = (E_{\text{long},i} + E_{\text{short},i}, Q_i), \quad (7)$$

where (\cdot) is the concatenation operator.

Thereafter, we use an inner product on the final embedding of the users and items to predict the recommendation results. The formula is as follows:

$$(\hat{y}_{u,i})_{|U| \times |I|} = E_u^* (E_i^*)^T. \quad (8)$$

3.5. Training. In this work, the Bayesian Personalized Ranking (BPR) loss [36] is used, which is a pairwise loss that encourages the prediction of an interacted entry to be higher than its uninteracted counterparts:

$$L_{BPR} = \sum_{(u,i,j) \in O} -\ln \sigma(\hat{y}_{u,i} - \hat{y}_{u,j}), \quad (9)$$

where $O = \{(u, i, j) | (u, i) \in O^+, (u, i) \in O^-\}$, is the dataset in the training process, which consists of interacted pairs set O^+ and uninteracted pairs set O^- . What is more, L_2 regularization is used to optimize the model parameter to prohibit overfitting risk. Therefore, the final objective function in our model is combined by BPR loss and regularization:

$$L_{our} = L_{BPR} + \gamma_1 \|\Theta_1\|_2^2 + \gamma_2 \|\Theta_2\|_2^2, \quad (10)$$

where set $\Theta_1 = \{E_u, E_i\}$ is the set of embedding parameters, $\Theta_2 = \{w_1, w_2, w'_1, w'_2\}$ is the set of weights in CNN layers, and γ_1, γ_2 are the hyperparameters to control the regularization. Furthermore, the Adam [37] is used in a minibatch manner to optimize the proposed model.

4. Experiments

Empirical results are proposed to evaluate the proposed model. The experiments aim to answer the following research questions:

RQ1: How does DSAGR perform as compared with state-of-the-art models?

RQ2: How do dynamic features affect DSAGR?

RQ3: What are the effects of hyperparameters on the DSAGR model?

4.1. Dataset. The dynamic preference learning module in the proposed method requires implicit feedback and temporal information; thus, the proposed model is evaluated on ML_100k and ML_1M movie datasets (<https://grouplens.org/datasets/movielens/>). Table 1 presents the statics of the datasets. These datasets have 5-level rating scores, and each user has rated at least 20 movies. The ratings of the datasets are binarized because the proposed model only requires implicit feedback. Specifically, every element in the original rating matrix (scores 1 to 5) is binarized to 1 and 0, where 1 indicates that the rating score is not less than 4, 0 indicates the rating score is less than 4, and no interaction. This work

TABLE 1: Statistics of the datasets.

Statistics	ML_100k	ML_1M
Number of users	944	6040
Number of items	1683	3952
Number of ratings	100000	1000209

also follows the same settings described in NGCF [5] to select 20% of interaction recodes randomly from each user to represent the test and valid sets and then treat the remaining as the training set.

4.2. Experimental Settings

4.2.1. *Baselines.* The GSAGR model is compared with the following methods:

- (i) Item-based CF (ICF) [38]: ICF is usually a two-step process: (1) determining the similarity set for target items and (2) predicting rating scores based on the most similar items. The rating scores of unseen items for the user are predicted in the second phase according to the weighted average rating of his k -nearest neighbor.
- (ii) PMF [1]: With the probabilistic matrix factorization (MF) algorithm, this model maps the user-item rating matrix into two low-dimensional matrixes. Then, this algorithm predicts the preference of users by the inner product between the two low-dimensional matrixes.
- (iii) DMF [39]: DMF is an MF-based CF method, which obtains the latent features of users and items through deep representation learning, that is, MLP. This method then uses the inner product between the two latent features to predict the preferences of users on items.
- (iv) Wide&Deep [17]: Wide&Deep is a famous deep learning recommender system that combines wide linear models and MLP neural network layers to obtain latent representations of users and items.
- (v) NGCF [5]: This work learns the representations of users and items by aggregating the information of high-order neighbors. Specifically, each node obtains the transformed representation of neighbors by propagating embeddings on the bipartite graph structure. NGCF introduces collaborative signals in the pooling layer to enhance high-order latent features learning.
- (vi) DGCF [40]: DGCF is an advanced graph-based CF model. This work focuses on the intentions of users for interacting with different items. The implementation of DGCF is based on the NGCF and graph attention network to model different intents of users.

4.2.2. *Evaluation Metrics.* Unlike the previous studies [17, 39] that perform metrics from sampled uninteracted items, this experiment conducts metrics for all the items that

the user has not interacted with. Two widely used evaluation protocols Recall@ N and NDCG@ N (normalized discounted cumulative gain) ($N = 20$ by default) are adopted to evaluate the effectiveness of top- N recommendation and preference ranking. The specific formula is as follows:

$$\text{Recall@}N = \frac{1}{|U|} \sum_u \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (11)$$

where TP (i.e., True Positive) indicates the number of items in the top- N recommendation list that hit the target items and FN (i.e., False Negative) is the number of the positive items in the test set that are falsely identified as the negative items.

$$N \text{ DC G@}N = \frac{1}{|U|} \sum_u \frac{DC \text{ G@}N}{I \text{ DC G@}N}, \quad (12)$$

where $DC \text{ G@}N = \sum_{i=1}^N 2^{r_i} - 1 / \log_2(i + 1)$; here, $r_i = 1$ if the test item is in position i , else 0; $I \text{ DC G@}N$ indicates the ideal $DC \text{ G@}N$ such that the target items are present at the top of the recommendation list.

4.2.3. *Parameter Settings.* The DSAGR model is implemented in Python under the TensorFlow (<https://www.tensorflow.org>) framework. For comparison algorithms, the parameter settings are given in the original works of literature. The proposed model uses the following parameter settings: (1) a random normal distribution (the mean and standard deviation are set to 0 and 0.01, resp.) is used to initiate the embedding matrix of users and items. Furthermore, the dimensionality of the embedding matrix is set to 64; (2) the delay factor for dynamic graph construction is set to one three-hundredth of the length of the dataset. (3) GCN and pooling layers with the hyperparameter $\lambda_0 = \lambda_1 = 1$ are used to represent the interaction features of users and items; (3) two GCN layers with 2 and 32 filter factors are used, and the kernel size in each layer is 3; (4) following NGCF [5] and DGCF [40], Adam optimization is used to train the model. The learning rate of the Adam algorithm is 0.0003, which is set by experiments.

4.3. Results and Discussion

4.3.1. *Performance Comparison (RQ1).* To answer the first research question, the proposed model is compared with six other methods in terms of Recall@ N and NDCG@ N . Two of the methods, ICF, and PMF are traditional and are frequently used CF algorithms. DMF and Wide&Deep are deep learning-based CF models. The remaining two, referred to as DGCF and NGCF, are versions of GCN with graph structure. Table 2 reports the performance for each algorithm. The following observations from this table are presented.

- (1) This table reveals that DSAGR has achieved the best result on ML_100k and ML_1M datasets. On metrics Recall@ N and NDCG@ N , DSAGR has improved the performance by at least 2.72% and 4.81%, respectively. For instance, DSAGR improves the NDCG@ N over the strongest baselines by 5.798% and 4.81% on

TABLE 2: Performance comparison of different methods.

Datasets					
Metrics	ML_100k		ML_1M		
	Recall@N	NDCG@N	Recall@N	NDCG@N	
ICF	0.1642	0.2913	0.1504	0.2030	
PMF	0.2303	0.3061	0.1943	0.2301	
DMF	0.3396	0.3562	0.2215	0.2472	
Wide&Deep	0.3402	0.3797	0.2502	0.2613	
NGCF	0.3487	0.4225	0.2637	0.2947	
DGCF	0.3339	0.4057	0.2973	0.3264	
Ours	0.3672	0.4470	0.3054	0.3421	

ML_100k and ML_1M, respectively. DSAGR can employ the dynamic information simply to provide additional side information for prediction by constructing the degree matrix. Meanwhile, DGCF and NGCF only aggregate the presentations of adjacent nodes in the graph. Significantly, DGCF uses multi-intent-aware graphs but performs worse than the proposed DSAGR model. The reason is that DGCF ignores the dynamic features and the short-term collaborative signals.

- (2) DGCF, NGCF, DMF, and Wide&Deep achieve better performance than traditional methods PMF and ICF. Therefore, compared with the CF only, the employment of deep learning and GCN is advantageous across the board. Wide&Deep and DMF fail to go beyond NDCG@N despite outperforming DGCF in Recall@N on the ML_100k dataset, implying that the graph-based methods are more effective than the deep learning methods in modeling the preference of users. In particular, the NGCF model improves Recall@N compared with the Wide&Deep model, with enhancements of 2.44%, 5.40%, on ML_100k and ML_1M datasets, respectively. The NGCF model also outperforms the DMF, with at least improvements of 2.68% and 18.61% on Recall@N and NDCG@N, respectively. Such improvement might be attributed to the GCN module, which captures more complex behavior patterns than MLP.

Furthermore, the experiment is repeated for all methods 10 times. Therefore, the freedom degree of t -distribution is 9. Specifically, this experiment accepts the hypothesis that DSAGR achieves better performance than baseline models on the two datasets for significance levels of 0.005. The statistical tests and results for this analysis are shown in Table 3. This table reveals that the method DSAGR successfully enhances the representation of users and items by considering the dynamic features and preferences.

4.3.2. Effect of the Proposed Technologies (RQ2). The proposed DSAGR is compared with different variants on the ML_100k and ML_1M datasets to investigate the superiority of the key technologies proposed in this work. Table 4 reports the variant models and their performances. The following findings are presented: DSAGR-L performs better

than DSAGR-S, which removes the long-term information. This finding is probably because the preference of users cannot be captured by the short-term information alone. DSAGR-DL, DSAGR-DG, and DSAGR also outperform DSAGR-D and DSAGR-G. This phenomenon proves that the captured dynamic features and short-term information can effectively improve the model’s performance. Moreover, DSAGR performs better than GRU-based [20] variant DSAGR-DG and LSTM-based [19] variant DSAGR-DL. This result is probably due to the small length of the row vectors of the dynamic matrix, allowing the CNN to model their dynamic features effectively.

4.3.3. The Sensitivity of Hyperparameters (RQ3). This work investigates how four hyperparameters, namely, the number of time slices, the filter factors in the first and second CNN layers, and the embedding size, affect DSAGR to examine the effect of the constructed dynamic graph among dynamic preference of users. The experiments on two datasets are conducted, providing similar rules, and only the results on the ML_100k are presented herein.

Inspired by the work [41], the experiment also adopts the orthogonal experimental design (OED) method to get a reasonable combination of these hyper-parameters. Specifically, the number of levels for the four parameters is set as follows: four levels for the time slices {11, 21, 31, 41}; four levels for the filter factors in the first CNN layer {2, 4, 6, 8}; four levels for the filter factors in the second CNN layer {8, 16, 32, 64}; and four levels for the embedding factors {16, 32, 64, 72}. A full-factorial analysis needs $4^4 = 256$ experiments. Taguchi’s method employs the orthogonal arrays to obtain the possible combinations of the hyperparameters from the whole combinations, thus bringing a minimum experimental run and the best estimation of parameters during the execution. In our experiments, the orthogonal array $L_{16}(4^4)$ has only 16 experiments, as shown in Table 5. This table shows that DSAGR can achieve better performance by setting time slices as 31, filter factors in the first and second CNN layers as 2 and 32, and embedding factors as 64.

The average values of Recall@N are used to investigate the effect of each factor. For example, the mean value of the first 4 rows in Table 5 is calculated to investigate the effect of time slice with level 11. The average values of Recall@N with different factors are shown in Figures 3–6.

TABLE 3: The t -test for paired comparisons in terms of Recall@N on the datasets.

ICF	PMF	DMF	Wide&Deep	NGCF	DGCF
<i>ML_100k</i>					
337.98	181.01	17.19	14.98	12.88	57.77
<0.005	<0.005	<0.005	<0.005	<0.005	<0.005
<i>ML_1M</i>					
305.99	141.38	225.11	73.34	67.96	14.95
<0.005	<0.005	<0.005	<0.005	<0.005	<0.005

TABLE 4: Performance of compared with different variants of DSAGR (“—” indicates DSAGR removes the key technology).

Variants	Dynamic preference		Dynamic feature	ML-100k		ML_1M	
	Long-term	Short-term		Recall@N	NDCG@N	Recall@N	NDCG@N
DSAGR-S	—	GCN	CNN	0.36013	0.44148	0.2891	0.3214
DSAGR-L	GCN	—	CNN	0.36370	0.32779	0.3000	0.3325
DSAGR-G	GCN	—	—	0.33217	0.4041	0.2345	0.2911
DSAGR-D	GCN	GCN	—	0.36006	0.44097	0.2798	0.3154
DSAGR-DL	GCN	GCN	LSTM	0.36184	0.44274	0.2921	0.3255
DSAGR-DG	GCN	GCN	GRU	0.36456	0.44921	0.2966	0.3310
DSAGR	GCN	GCN	CNN	0.3672	0.4470	0.3054	0.3421

TABLE 5: Performance of 16 experiments obtained from Taguchi’s method.

ID	Time slices	Filter-first CNN layer	Filter-second CNN layer	Embedding factors	Recall@N	NDCG@N
1	11	2	8	8	0.3422	0.4200
2	11	4	16	16	0.3552	0.4369
3	11	6	32	32	0.3617	0.4420
4	11	8	64	64	0.3540	0.4296
5	21	2	16	32	0.3560	0.4373
6	21	4	8	64	0.3626	0.4412
7	21	6	64	8	0.3451	0.4224
8	21	8	32	16	0.3438	0.4186
9	31	2	32	64	0.3672	0.4470
10	31	4	64	32	0.3660	0.4481
11	31	6	8	16	0.3463	0.4237
12	31	8	16	8	0.3469	0.4259
13	41	2	64	16	0.3547	0.4232
14	41	4	32	8	0.3483	0.41786
15	41	6	16	64	0.3533	0.4217
16	41	8	8	32	0.3491	0.4286

- (1) *Effect of Time Slice Numbers.* The number of time slices determines the number of dynamic subgraphs and the last subgraph. As shown in Figure 3, DSAGR can achieve the best performance by setting the time slice as 31. Figure 3 shows that when the number of time slices reaches 31, adding more time slices cannot improve the recommendation performance. Also, more time slices will increase the dimension of the row vectors of the degree matrix, and consequently, there will be an increase in the training time taken.
- (2) *Effect of Filter Factors in CNN.* Figures 4 and 5 show the recommendation performance of different filters in the first and second CNN layers. Figure 5 reveals

that the performance gradually becomes better with the increase of filter factors in the second CNN layer. However, blindly increasing the filter factors does not necessarily improve the performance of DSAGR. This is maybe because that more information is encoded when the filter factors become larger, but it may also bring a little overfitting.

- (3) *Effect of Embedding Factors.* Figure 6 illustrates the performance of DSAGR under the different embedding factors. The figure reveals that the performance of the model gradually improves as the dimensionality increases. And the performance tends to be stable when the embedding factors are set as 64.

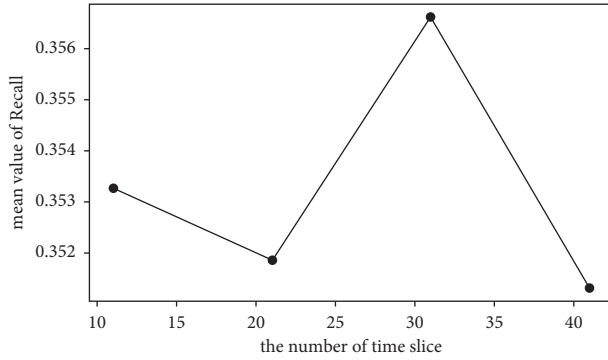


FIGURE 3: Effect of time slice numbers.

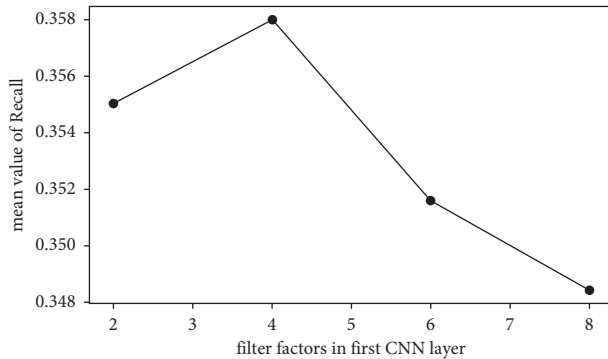


FIGURE 4: Effect of filter factors in the first CNN layer.

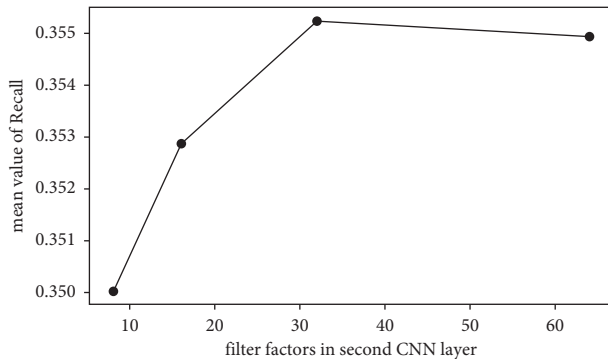


FIGURE 5: Effect of filter factors in the second CNN layer.

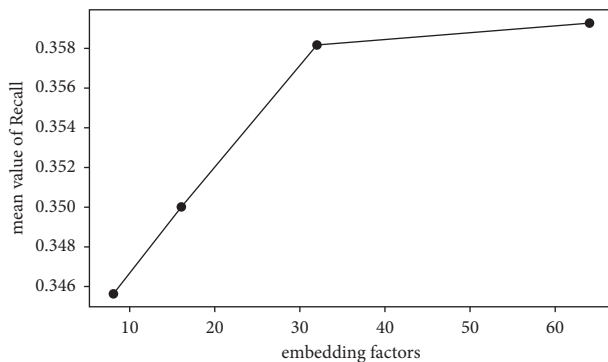


FIGURE 6: Effect of embedding factors.

5. Conclusion

In this work, a hybrid recommender system is proposed to capture the dynamic preferences of users and dynamic sequence features. The proposed model, namely, DSAGR, combines GCN and CNN to obtain the latent representations of users and items and then makes a prediction. The dynamic preference is modeled by the long- and short-term interaction graphs of users. The dynamic sequence includes the degree matrixes of users and items captured from the dynamic graph. To our knowledge, this type of modeling using the short-term interaction graph and degree matrixes has not been previously applied to predict users' preferences. The experimental results show that the DSAGR model significantly improves the performance compared with baselines. Considering future work, two feasible avenues are available: (1) The work concentrates on learning the latent representation of users and items via dynamic information. Thus, one direction of further study is to design an effective way to aggregate the long- and short-term representations to a single vector, which is successful in maximizing deep learning. (2) The method of capturing dynamic features provides a new idea to many other unstructured data, such as social networks. It is worth trying to improve the recommendation performance.

Data Availability

The data used include ML_100k and ML_1M movie datasets. The movie dataset address is as follows: <https://grouplens.org/datasets/movielens/>.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Tao Chen and Ninghua Sun conceived and designed the experiments; Ninghua Sun proposed the method and performed the experiments; Longya Ran analyzed the data; Ninghua Sun prepared the original draft; and Longya Ran and Wenshan Guo reviewed and edited the manuscript. All authors have read and agreed on the published version of the manuscript.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities, China (HUST: 2020JYCXJJ036), Humanities and Social Science Fund of the Ministry of Education of China (19YJA630010), National Natural Science Foundation of China (71734002 and 72042016), and Key R&D Projects, Hubei Province (2021BAA033).

References

- [1] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Proceedings of the 20th International Conference on Neural Information Processing Systems*, pp. 1257–1264, Red Hook, NY, USA, 2007.
- [2] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. S. Chua, "Neural collaborative filtering," in *Proceedings of the 26th International World Wide Web Conference, WWW 2017, International World Wide Web Conferences Steering Committee*, pp. 173–182, Republic and Canton of Geneva, Geneva, Switzerland, 2017.
- [3] Y. Guo and Z. Yan, "Recommended system: attentive neural collaborative filtering," *IEEE Access*, vol. 8, 2020.
- [4] H. Liu, Y. Wang, Q. Peng et al., "Hybrid neural recommendation with joint deep representation learning of ratings and reviews," *Neurocomputing*, vol. 374, pp. 77–85, 2020.
- [5] X. Wang, X. He, M. Wang, F. Feng, and T.-S. Chua, "Neural graph collaborative filtering," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 165–174, ACM, New York, NY, USA, 2019.
- [6] X. He, K. Deng, X. Wang, Y. Li, Y. Zhang, and M. Wang, "LightGCN: simplifying and powering graph convolution network for recommendation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 639–648, ACM, New York, NY, USA, 2020.
- [7] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021.
- [8] A. Aji, Y. Wang, E. Agichtein, and E. Gabrilovich, "Using the past to score the present," in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management - CIKM '10*, p. 629, 2010.
- [9] H. Takemura and K. Tajima, "Tweet classification based on their lifetime duration," in *Proceedings of the 21st ACM International Conference on Information and Knowledge Management - CIKM '12*, p. 2367, 2012.
- [10] M. Zhang, S. Wu, X. Yu, Q. Liu, and L. Wang, "Dynamic graph neural networks for sequential recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, p. 1, 2022.
- [11] J. C. Robinson, "The taken time-delay embedding theorem," in *Proceedings of the Dimensions, Embeddings, and Attractors*, pp. 145–159, Cambridge University Press, Cambridge, 2010.
- [12] H. Yu, L. T. Yang, Q. Zhang, D. Armstrong, and M. J. Deen, "Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives," *Neurocomputing*, vol. 444, 2021.
- [13] S. Lawrence, C. L. Giles, and A. D. Ah Chung Tsoi, "Back, Face recognition: a convolutional neural-network approach," *IEEE Transactions on Neural Networks*, vol. 8, 1997.
- [14] D. Wang, Y. Yih, and M. Ventresca, "Improving neighborhood collaborative filtering by using a hybrid similarity measurement," *Expert Systems with Applications*, vol. 160, p. 160, 2020.
- [15] Z. Hu, Y. Zhang, Y. Xing, Y. Zhao, D. Cao, and C. Lv, "Toward human-centered automated driving: a novel spatiotemporal vision transformer-enabled head tracker," *IEEE Vehicular Technology Magazine*, 2022.
- [16] Z. Hu, C. Lv, P. Hang, C. Huang, and Y. Xing, "Data-driven estimation of driver attention using calibration-free eye gaze and scene features," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 2, pp. 1800–1808, 2022.
- [17] H.-T. Cheng, L. Koc, J. Harmsen et al., "Wide & deep learning for recommender systems," in *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, 2016.
- [18] H. Wang, X. Shi, and D.-Y. Yeung, "Collaborative recurrent autoencoder: recommend while learning to fill in the blanks," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2016.
- [19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [20] K. Cho, B. van Merriënboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, Association for Computational Linguistics, Stroudsburg, PA, USA, 2014.
- [21] K. Sun, T. Qian, T. Chen, Y. Liang, Q. V. H. Nguyen, and H. Yin, "Where to go next: modeling long- and short-term user preferences for point-of-interest recommendation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 214–221, 2020.
- [22] L. Zheng, V. Noroozi, and P. S. Yu, "Joint deep modeling of users and items using reviews for recommendation," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 425–434, ACM, New York, NY, USA, 2017.
- [23] H. Wu, Z. Zhang, K. Yue, B. Zhang, J. He, and L. Sun, "Dual-regularized matrix factorization with deep neural networks for recommender systems," *Knowledge-Based Systems*, vol. 145, pp. 46–58, 2018.
- [24] C. Gao, X. Wang, X. He, and Y. Li, "Graph neural networks for recommender system," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pp. 1623–1625, ACM, New York, NY, USA, 2022.
- [25] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: a comprehensive review," *Computational Social Networks*, vol. 6, no. 1, p. 11, 2019.
- [26] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems*, I. Guyon, U. v Luxburg, S. Bengio et al., Eds., pp. 1025–1035, Curran Associates Inc., Red Hook, NY, USA, 2017.
- [27] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" in *Proceedings of the International Conference on Learning Representations*, New Orleans, LA, USA, 2019.
- [28] J. Li, H. Dani, X. Hu, J. Tang, Y. Chang, and H. Liu, "Attributed network embedding for learning in a dynamic environment," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 387–396, ACM, New York, NY, USA, 2017.
- [29] R. Ye, Y. Hou, T. Lei et al., "Dynamic graph construction for improving diversity of recommendation," in *Proceedings of the Fifteenth ACM Conference on Recommender Systems*, pp. 651–655, ACM, New York, NY, USA, 2021.
- [30] H. Xu, C. Huang, Y. Xu, L. Xia, H. Xing, and D. Yin, "Global context enhanced social recommendation with hierarchical graph neural networks," in *Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM)*, pp. 701–710, IEEE, Sorrento, Italy, 2020.
- [31] W. Song, Z. Xiao, Y. Wang, L. Charlin, M. Zhang, and J. Tang, "Session-based social recommendation via dynamic graph attention networks," in *Proceedings of the Twelfth ACM*

- International Conference on Web Search and Data Mining*, pp. 555–563, ACM, New York, NY, USA, 2019.
- [32] Q. Liu, Y. Zeng, R. Mokhosi, and H. Zhang, “STAMP : short-term attention/memory priority model for session-based recommendation,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1831–1839, ACM, New York, NY, USA, 2018.
- [33] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, “Signature verification using a “siamese” time delay neural network,” in *Proceedings of the 6th International Conference on Neural Information Processing Systems*, pp. 737–744, San Francisco, CA, USA, 1993.
- [34] J. Son and S. B. Kim, “Academic paper recommender system using multilevel simultaneous citation networks,” *Decision Support Systems*, vol. 105, pp. 24–33, 2018.
- [35] T. Pradhan and S. Pal, “A hybrid personalized scholarly venue recommender system integrating social network analysis and contextual similarity,” *Future Generation Computer Systems*, vol. 110, pp. 1139–1166, 2020.
- [36] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, “BPR: bayesian personalized ranking from implicit feedback,” in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, pp. 452–461, AUAI Press, Arlington, Virginia, USA, 2009.
- [37] D. P. Kingma and J. Ba, “Adam: a method for stochastic 7 optimization,” in *Proceedings of the International Conference on Learning Representations (ICLR)*, Ithaca, NY, USA, May 2015.
- [38] G. Karypis, J. Konstan, and J. Riedl, “Item-based collaborative filtering recommendation algorithms,” in *Proceedings of the 10th International Conference on World Wide Web*, pp. 285–295, New York, NY, USA, 2001.
- [39] H.-J. Xue, X. Dai, J. Zhang, S. Huang, and J. Chen, “Deep matrix factorization models for recommender systems,” in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 3203–3209, International Joint Conferences on Artificial Intelligence Organization, California, 2017.
- [40] X. Wang, H. Jin, A. Zhang, X. He, T. Xu, and T.-S. Chua, “Disentangled graph collaborative filtering,” in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020.
- [41] S. Gao, M. Zhou, Y. Wang, J. Cheng, H. Yachi, and J. Wang, “Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 2, pp. 601–614, 2019.

Research Article

Computational Methods for Automated Analysis of Malaria Parasite Using Blood Smear Images: Recent Advances

Shankar Shambhu ¹, Deepika Koundal ², Prasenjit Das ¹, Vinh Truong Hoang ³,
Kiet Tran-Trung ³ and Hamza Turabieh ⁴

¹Chitkara University School of Computer Applications, Chitkara University, Himachal Pradesh, India

²School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India

³Faculty of Computer Science, Ho Chi Minh City Open University, Ho Chi Minh City, Vietnam

⁴Department of Information Technology, College of Computing and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

Correspondence should be addressed to Deepika Koundal; dkoundal@ddn.upes.ac.in

Received 14 January 2022; Accepted 26 March 2022; Published 11 April 2022

Academic Editor: Zhongxu Hu

Copyright © 2022 Shankar Shambhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Malaria comes under one of the dangerous diseases in many countries. It is the primary reason for most of the causalities across the world. It is presently rated as a significant cause of the high mortality rate worldwide compared with other diseases that can be reduced significantly by its earlier detection. Therefore, to facilitate the early detection/diagnosis of malaria to reduce the mortality rate, an automated computational method is required with a high accuracy rate. This study is a solid starting point for researchers who want to look into automated blood smear analysis to detect malaria. In this paper, a comprehensive review of different computer-assisted techniques has been outlined as follows: (i) acquisition of image dataset, (ii) preprocessing, (iii) segmentation of RBC, and (iv) feature extraction and selection, and (v) classification for the detection of malaria parasites using blood smear images. This study will be helpful for: (i) researchers can inspect and improve the existing computational methods for early diagnosis of malaria with a high accuracy rate that may further reduce the interobserver and intra-observer variations; (ii) microbiologists to take the second opinion from the automated computational methods for effective diagnosis of malaria; and (iii) finally, several issues remain addressed, and future work has also been discussed in this work.

1. Introduction

Malaria has turned into a major risk to individuals worldwide as one of the main reasons for causalities across the world. It is a curable infectious disease caused by a protozoan parasite that can be life-threatening. As per the latest report of the World Health Organization (WHO), in 2019, 229 million malaria cases were detected worldwide, and causalities were reached to 409000. In 2018, 228 million malaria cases were detected, and causalities were reached 411000 [1].

In 2016 and 2017, about 1.09 million and 0.84 million malaria cases were registered in India, in which most of the malaria cases were *P. falciparum* species affected [2].

Dr. Ronald Ross first discovered malaria transmission in the human body by mosquitoes in 1897 [3]. The main reason for malaria is a protozoan parasite. The plasmodium genus infects the red blood cells (RBC) of the human body, which causes malaria [4]. In general, female *Anopheles* mosquitoes and human beings are the two main hosts infected by the parasite. When female *Anopheles* mosquitoes desire to foster their eggs, they bite and draw blood from the human body. If a parasite infects that person, then that same infected parasite blood is found in the mosquito and that parasite reproduces and develops in the mosquito body. When that infected mosquito bites another person, parasites containing the salivary gland are transferred into that person's blood [5]. After transferring parasites into the human body by the

mosquito, malaria parasites grow with very high speed in the liver and RBC of that infected person. Symptoms of malaria appear after one or two weeks. Primary symptoms that appear are headache, vomiting, fever, and chills. If malaria is not treated early and properly, it is very harmful to the human body. It may be a reason for kidney failure, low blood sugar, respiratory distress, enlargement of the spleen, etc. [6]. Malaria can kill a person by destroying their RBC. Malaria during pregnancy is very dangerous, and it is one of the reasons for abortion [7].

There are five different protozoan parasite species, which are the main cause of malaria in the human body. These are *Plasmodium falciparum* (*P. falciparum*), *Plasmodium vivax* (*P. vivax*), *Plasmodium ovale* (*P. ovale*), *Plasmodium malariae* (*P. malariae*), and *Plasmodium knowlesi* (*P. knowlesi*). Among all five species, the first four are the most common species, which occur in the human body. The fifth species is *P. knowlesi* mostly occurs in monkeys that live in South-East Asia forests. But, in past years, some cases of *P. knowlesi* malaria occurred in the human body. The most common species found in the human body is *P. vivax*, but the most dangerous species is *P. falciparum* [8]. Figure 1 shows the images of the different types of malaria found in human peripheral blood smears.

All species of protozoan parasites are morphologically different. At every stage of its lifecycle, each species changes in its size, color, shape, and morphology. These various stages of every species are ring, trophozoite, schizont, and gametocyte, as shown in Figure 2.

The main reason for the high mortality rate is the late detection of malaria. In medical science, for the detection of malaria, microscopic examination is the gold standard. A microbiologist manually counts affected RBC under the microscope to examine the patient's blood sample, which is a very time-consuming and highly tedious process. The accuracy of this process is entirely dependent on microbiologist expertise [10]. Hence, microscopic examination is a prolonged process, and it is the main reason for the late detection of malaria in patients, increasing the high mortality rate. The high malaria mortality rate can be decreased by detecting malaria at an early stage. Therefore, an automated computer-assisted technique is needed, which will help the microbiologists to provide a second opinion for effective and early detection of malaria and reduce the mortality rate.

The pattern of total worldwide malaria patients is illustrated in Figure 3. It represents how malaria patients are increasing worldwide. In 2013, 198 million malaria-affected patients were detected, which was increased to 229 million in 2019 [1]. These very troubling statistics can be reduced by detecting parasites and diagnosis in the early stages, and it would be beneficial when experts are not available.

The paper's contributions are as follows: (i) a comprehensive review has been conducted on the state-of-the-art techniques for malaria diagnosis that have been published in the last decade; (ii) various types of automated computational methods such as preprocessing, segmentation, feature extraction, and classification for diagnosing malaria have been discussed in detail; (iii) additionally, different types of

machine learning and deep learning models, as well as their accuracies for malaria parasite detection and diagnosis, have been discussed; (iv) moreover, several types of blood smear image datasets for malaria diagnosis have been identified; and (v) various challenges and issues with the already implemented techniques and scope of future work have also been discussed.

The paper is organized as follows: (i) Section 2 summarizes the state-of-the-art techniques for malaria diagnosis; (ii) Section 3 explains automated computational methods for diagnosing malaria in detail; (iii) Section 4 presents the discussion with research gaps; and (iv) Section 5 concludes the paper with future scope.

2. State-of-the-Art Techniques for Malaria Diagnosis

Malaria is a disease in which symptoms appear after 7 to 15 days. Primary symptoms are headache, vomiting, fever, pain, chills, etc. These symptoms could be an indication of malaria, although many diseases have the same symptoms. Hence, some techniques are needed that can diagnose malaria correctly. For malaria diagnosis, different techniques have been developed such as microscopy blood smear examination, cytometry, rapid diagnostic test (RDT), polymerase chain reaction, and fluorescent microscopy. Still, for diagnosing malaria, the primarily used techniques are (a) microscopic thick and thin blood smears examination and (b) rapid diagnosis test in medical science [11].

2.1. Microscopic Thick and Thin Blood Smears Examination. In this, a laboratory examination is performed in which a clinician divides the blood sample into two parts on the slide. One is called a thick blood smear, and another is a thin blood smear. After that, a clinician manually counts the affected RBC under the microscope. A thick blood smear helps clinicians detect the presence of malaria parasites, and a thin blood smear helps identify the species of the parasites causing malaria. All the steps for malaria detection using microscopic blood smears examination are shown in Figure 4.

Advantages of the microscopic technique are as follows: (i) a clinician can distinguish the different stages of malaria species at a very low cost using microscopic method and (ii) microscopy technique for malaria detection is more effective as compared to rapid diagnostic tests as it can count affected RBC very efficiently. Apart from the advantages of microscopic techniques for malaria detection, some challenges are also there. Microscopic thick and thin blood smears examinations technique accuracy depends on microbiologist experience. To detect and diagnose malaria through a microscope, a microbiologist may have to count malaria-affected RBC manually, which is a highly tedious and time-consuming task [10]. It is found in multiple studies that manual counting of affected cells using a microscope is not an authentic technique when it is done by a nonexperienced microbiologist [13]. Instead of this, to confirm a blood smear slide is malaria-affected or not, a microbiologist needs

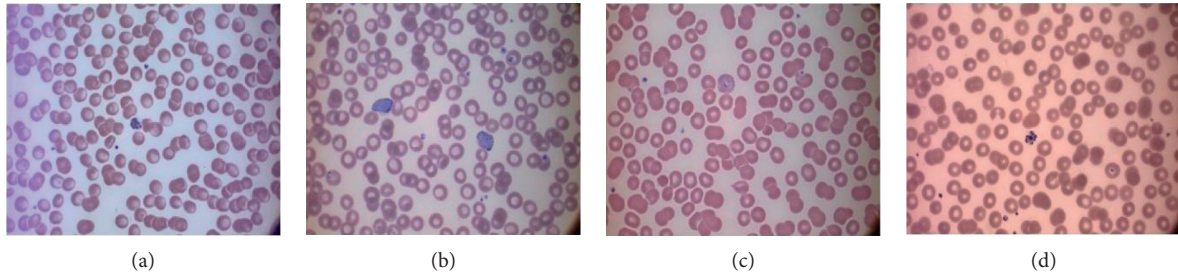


FIGURE 1: Different types of malaria peripheral blood smear images (a) *P. falciparum* (b) *P. vivax* (c) *P. ovale* (d) *P. malaria* [9].

Species	STAGES			
	Ring	Trophozoite	Schizont	Gametocyte
<i>P.falciparum</i>				
<i>P.vivax</i>				
<i>P.ovale</i>				
<i>P.malaria</i>				

FIGURE 2: Different stages of malaria parasite species.

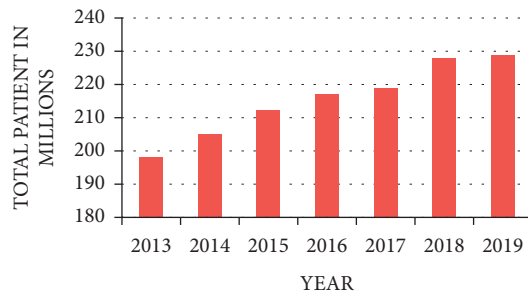


FIGURE 3: Worldwide year-wise prevalence count of malaria patients.

significant time. But, it is a tough task for a microbiologist to examine each slide because a microbiologist has to study multiple blood smear images under the microscope. Moreover, this technique takes time to examine blood smear slides.

2.2. Rapid Diagnosis Test (RDT). Rapid diagnosis test or antigen test is a small kit used to detect antigens derived from malaria parasites. To identify malaria, a drop of blood is inserted into the kit from the given hole, and internally, this

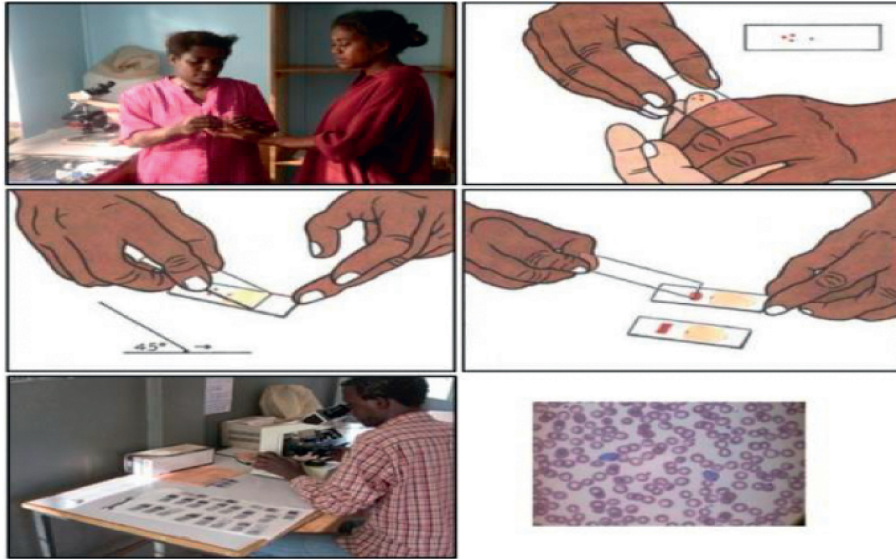


FIGURE 4: Microscopic thick and thin blood smears examination [12].

device performs the tests and provides the result in minimum time. RDT kit functioning is shown in Figure 5.

Advantages of the RDT kit are as follows: (i) it is significantly faster than manual cell counting techniques, and it gives instant results; (ii) for the use of the RDT kit, no expertise is required; and (iii) it is beneficial in endemic regions. Instead of the advantages of the RDT kit for malaria detection, some challenges are also there. As per the analysis of different studies, the results of this technique are less accurate, and any wrong result can affect the patient's treatment [14]. Another main challenge of the RDT kit is detecting whether a patient is malaria-affected or not. It cannot detect malaria species.

Hence, after studying different techniques of malaria diagnoses and their advantages and challenges, researchers observed that a computer-assisted malaria detection technique would be required. A computer-assisted malaria detection technique increases the performance of existing techniques by avoiding its limitations in terms of accuracy, instant results, dependency, and requirement of the expert microbiologist.

3. Automated Computational Methods for Diagnosis of Malaria

In medical science, the computer plays a very crucial role. Different automated computational methods are used for the diagnosis of multiple diseases. Ultrasound images, magnetic resonance imaging, X-ray images, and computed tomography images are used to diagnose different diseases of human anatomy using computerized imaging techniques. The computer-assisted diagnosis technique for malaria is based on the microscopic technique, which is performed by computer with the help of machine learning algorithms and computer vision techniques. This is the technique in which digital thin and thick blood smear images are used for the detection of malaria parasites automatically. Different steps of automated diagnosis of malaria are image acquisition, preprocessing, red blood cell

detection and segmentation, feature extraction, and selection and classification (parasite identification and labeling). The stepwise process of automated computational methods for malaria parasite diagnosis is shown in Figure 6. In this section, a deep survey has been performed on each technique used for automated detection of malaria using blood smear images.

3.1. Acquisition of Image Dataset. Digital images of blood smear samples are required to detect malaria in a patient using computer vision image processing and machine learning techniques. Each patient's blood smear sample is distributed into two parts: thick and thin blood smear images. Most computer-assisted detection studies use thin blood smear digital images, and very few researchers have worked on thick digital blood smear images [16].

Figure 7 shows the images of thick and thin blood smears. A thick blood smear is a drop of blood that assists in detecting the presence of parasites, and a thin blood smear is a layer of blood that is spread on a glass slide and assists in identifying the species of the parasite causing the infection. Different sources collect digital blood smear images, and this process is called the image acquisition technique. Categorization of different image acquisition techniques used on blood smear images for malaria parasite detection is shown in Table 1.

After analyzing the different image acquisition techniques in Table 1, we observed that there are various image acquisition techniques available. Still, the light microscopy technique is the most widely used and preferred because it has a high magnification factor, and it is beneficial for viewing the surface details of a blood smear.

Furthermore, Table 2 lists the different datasets of light microscopy techniques used by various researchers.

3.2. Preprocessing. Preprocessing is a technique used to remove the unwanted noise and produce high contrast digital blood smear images for the next step. When different

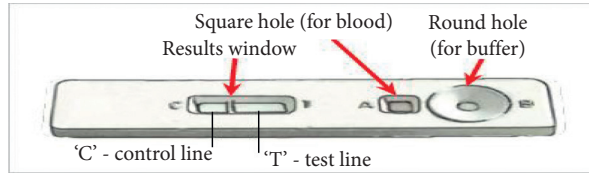


FIGURE 5: Rapid diagnosis testing (RDT) kit [15].

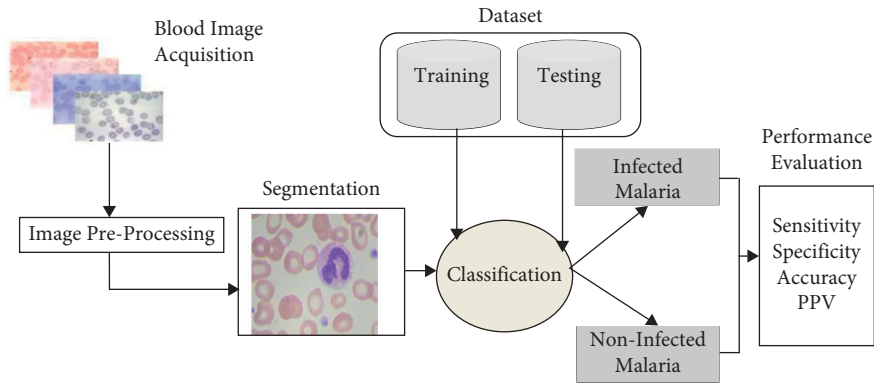


FIGURE 6: Computational methods for automated diagnosis system for malaria.

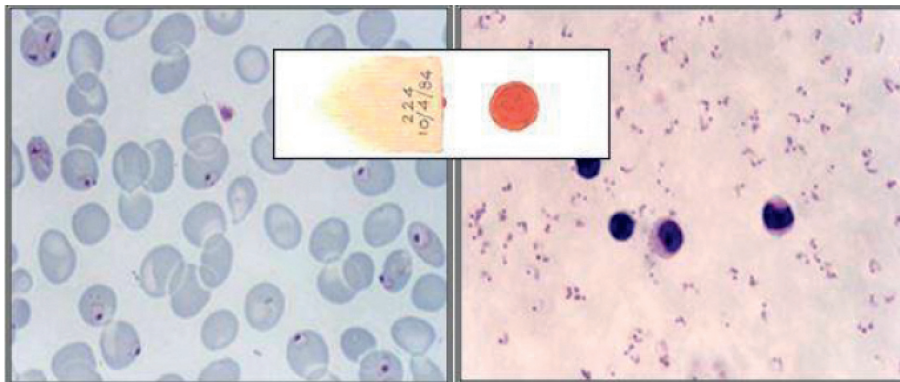


FIGURE 7: Malaria infected thin (left) and thick (right) blood smear image.

TABLE 1: Categorization of image acquisition techniques used on blood smear images for malaria parasite detection.

References	Light microscopy	Binocular microscopy	Fluorescent microscopy	Polarized microscopy	Multispectral and multimodal microscopy	Image-based cytometer	Scanning electron microscopy
[17]	✓						
[18]			✓				
[19]	✓						
[20]	✓						
[21]	✓						
[22]			✓				
[23]	✓						
[24]						✓	
[25]	✓						
[26]	✓						
[27]		✓					
[28]	✓						
[29]	✓						
[30]				✓			

TABLE 1: Continued.

References	Light microscopy	Binocular microscopy	Fluorescent microscopy	Polarized microscopy	Multispectral and multimodal microscopy	Image-based cytometer	Scanning electron microscopy
[31]					✓		
[32]	✓						
[33]	✓						
[34]			✓				
[35]							✓
[36]	✓						
[37]					✓		
[38]		✓					
[39]	✓						
[40]	✓						
[41]		✓					
[42]			✓				

TABLE 2: Light microscopy datasets used by different researchers.

Reference	No. of images in dataset	Remarks
[43]	300 images	Used KNN classifier on light microscopic images, got 91% accuracy.
[33]	21 images	Light microscopic images of 1296×1024 resolution captured by an axiocam high-resolution color camera were used.
[44]	—	—
[29]	68 images	Light microscopic images of different magnification have been used.
[25]	300 images	Used KNN classifier and got 90.17% accuracy to detect malaria parasite species.
[26]	27578 images	27578 single cell light microscopy images were used, and a new 16-layer CNN model was proposed to identify malaria-infected or infected images.
[23]	160 images	Achieve 95% accuracy for the detection of malaria.
[45]	—	Used Giemsa-stained blood smear images were taken by a camera attached with a microscope on 1000x magnification, and the proposed model got 77.78% accuracy.
[19]	27558 images	Implement novel stacked convolutional neural network technique for parasite detection.

—Not reported by the original paper.

resources take blood smear images, the images are corrupted by noise, and thus, visualization of the images is not good. Due to this problem, further steps of segmentation and classification are challenging to implement, and it produces poor results. Hence, certain preprocessing techniques have been used to remove that unwanted noise from images. Preprocessing techniques remove the noise from the image for better visualization, which is very useful for further analysis [46]. As shown in Table 3, researchers used multiple preprocessing techniques such as median filter, mean filter, low-pass filter, morphological filter, partial contrast stretching, local histogram equalization, Laplacian filter, SUSAN filter, geometric mean filter, Gaussian filters, and Wiener filter for enhancing the contrast and remove the unwanted noise of digital images. May et al. have given an approach in which the median filter technique for removing the impulse noise from digital images and for removing additive noise has been used [4]. The Gaussian filter is used by Arco et al. to enhance the quality of the images affected by Gaussian noise. The geometric mean filter is also used by Das et al. for preserving the edges and removing Gaussian noise from the digital microscopic image [66]. Laplacian filter is used by Savkare et al. for smoothening and enhancing edges of malaria parasite images [29]. To preserve the structure of an image, Susan’s filter is suggested by [41]. To remove the

intensity of high frequency from a digital image, a low pass filter has been suggested by [55]. The histogram matching technique is used by Abbas et al. to normalize the intensity value of digital image pixels [67]. The categorization of preprocessing techniques to enhance the quality of digital blood smear images is shown in Table 3.

Table 4 displays the image preprocessing approaches used by various researchers for better visualization of thin and thick blood smear images with their properties.

3.3. Red Blood Cell Detection and Segmentation. Segmentation is the process in which digital images are disjoint into nonoverlapping regions. Each disjoint image typically corresponds to other parts of an image object. Once each digital image object is isolated, each object can be easily measured and classified. In the literature, different segmentation techniques have been applied on digital blood smear images to detect ROI (region of interest).

Das et al. have developed an automated system for classifying malaria at different stages. The researcher used the watershed segmentation technique in their research work for the segmentation of thin blood smear digital images. This technique provided better results for detecting erythrocytes from the whole blood smear image [66]. Further, a watershed algorithm is suggested by Savkare et al.

TABLE 3: Categorization of preprocessing techniques used on blood smear images.

References	MMF	LPF	MF	PCS	LHE	LF	SF	GMF	GF	WF
[47]	✓									
[48]										✓
[49]							✓			
[50]	✓		✓							
[29]	✓					✓				
[51]								✓		
[52]						✓				
[53]				✓						
[54]					✓					
[55]		✓								
[21]			✓							
[27]	✓									
[20]	✓									
[56]									✓	
[38]							✓			
[57]					✓					
[43]	✓									
[58]	✓									
[59]			✓							
[60]			✓							
[61]					✓					
[62]					✓				✓	
[63]	✓									
[64]			✓							
[24]	✓									
[65]			✓							

MMF—median/mean filter; LPF—low pass filter; MF—morphological filter; PCS—partial contrast stretching; LHE—local histogram equalization; LF—Laplacian filter; SF—SUSAN filter; GMF—geometric mean filter; GF—Gaussian filters; WF—Wiener filter.

to find overlapped cells on connected components [29]. Soni et al. used the granulometry technique [41]. Makkapati and Rao have developed a technique to segment RBC and parasites using HSV (hue, saturation, and value) color space. This technique segmented the RBC and parasites from the blood smear image based on hue range and optimal saturation thresholds [69]. Mandal et al. introduced a normalized cut method for microscopic blood smear images segmentation.

The segmentation algorithm has been used on different color spaces to find the optimal performance of digital blood smear images [70]. The result of the normalized cut segmentation algorithm is good in HSV color space. Nasir et al. have presented a segmentation-based approach using a K-means clustering algorithm for the segmentation of malaria parasite on 100 digital blood smear images dataset [71]. Bhatia also proposed a K-means clustering technique using genetic methods [72]. Panchbhai et al. have reported the RGB color space model and Otsu algorithm for RBC and parasite segmentation from 20 thin blood smear images [73]. For microscopic blood cells, digital images segmentation, the K-means clustering technique, and global threshold techniques are suggested by Savkare and Narote [74]. In this, 78 microscopic blood cell digital images are used for segmentation. Khan et al. also used the K-means clustering for the segmentation of 118 blood smear images to identify malaria parasite tissues [75]. Acharya et al. introduced a new computer-assisted detection technique for segmenting blood smear images and determining the acute myeloid leukemia

stage (AML). This work's approach is divided into many stages. A unique algorithm is being developed to accurately segment blood smear images in order to identify AML and its variants. The classification accuracy of the model was 99.48% on 500 test images [76].

To detect the exact radius of RBC, the circle hough transformation method was introduced by Ma et al. [77]. Otsu thresholding clustering-based method is presented by Makkapati et al. to get the image mask of binary image [78].

Deep learning techniques are also beneficial in image segmentation. Researchers for image segmentation have proposed many deep learning techniques. For image segmentation, a fully convolutional neural network-based deep learning technique has been proposed by Long et al. [79] and Wang et al. [80]. A completely CNN encoder and decoder deep learning segmentation technique (SegNet) has been used by Badriinarayanan et al. [81]. Ronneberger et al. proposed the U-Net to segment biomedical microscopic images [82]. Dai et al. created a multifunction network, for instance segmentation that includes three networks for separating instances, computing masks, and labeling objects. These networks must share their convolutional characteristics and form a cascaded structure [83]. Visin et al. have used ReSeg, an RNN-based deep learning approach for semantic segmentation of the images. This approach is primarily based on the image classification model ResNet [84].

Segmentation techniques on blood smear images used in different studies are summarized in Table 5. After analyzing

TABLE 4: Thin and thick blood smear based preprocessing techniques used for better visualization.

Type of blood smear	Problems	Reference	Preprocessing technique used	Remarks	Limitations/challenges
Thin blood smear image	Noisy blood smear image	[20, 27]	Median/Mean filter	Used to remove noise from blood smear images without affecting the edges.	The presence of impulse noise cannot be eliminated. It impacts the average rating of all pixels in the surrounding area.
		[48]	Wiener filter	Used to enhance the quality of blurred images.	The power spectra are difficult to estimate.
		[38, 49]	SUSAN filter	Helpful for finding the edges corners and for noise removal.	The brightness similarity metric is significantly affected by the threshold.
		[55]	Gaussian low-pass filter	For removing Gaussian noise in blood smear images, Gaussian low-pass filter was used.	Take too much time.
		[51]	Geometric mean filter	Useable for maintaining edges while removing Gaussian noise.	A negative observation will result in an imaginary geometric mean value regardless of the other observations' quantity.
	Low contrast blood smear image	[21, 50, 59]	Morphological filtering	Helpful for deleting unwanted objects, filling small holes, and splitting images.	When using morphological operators, it is necessary to consider the concepts of infimum and supremum.
		[53]	Partial contrast stretching method	Used to increase the contrast of the blood smear image.	—
		[29, 52]	Laplacian filter	Helpful for detection and improving the edges of the blood smear image.	The detection of edges and their directions increases the noise in the image, reducing the edge magnitude.
		[54, 57]	Local histogram equalization	Used to increase the resolution of blood smear images.	It is an indiscriminate technique.
		[62]	Low-pass filter	For removing excessive frequency components from blood smear images.	—
Thick blood smear image	Variations in cell staining	[20]	Gray world color normalization	Used for equality of color in blood smear images.	Poorly constructed normalization software might result in a reduction in the entire image quality.
			Gaussian low-pass filter		Take too much time.
		Laplacian filter		The detection of edges and their directions increases the noise in the image, reducing the edge magnitude	
	Noisy blood smear image	[68]	Median filter Local histogram equalization Contrast enhancement method		

—Not reported by the original paper.

various segmentation techniques, it was found that for the segmentation of malaria parasites and RBC, most researchers used Watershed, Marker-controlled watershed, and Edge detection algorithm, and deep learning techniques at the segmentation phase. For the segmentation of overlapping cells, watershed algorithm results are best [28].

3.4. Feature Extraction and Selection. Feature extraction after segmentation is a prerequisite for feature selection and classification. The objective of feature extraction is to recognize and characterize an object whose dimensions are very nearest or similar for objects in the same class and different for objects from a different class. It reduces the

TABLE 5: Summary of segmentation techniques used on digital blood smear images.

References	Segmentation techniques used	Remarks	Limitations/Challenges
[4, 29, 43, 85–88]	Otsu thresholding	Classification of pixels by using a calculating optimum threshold value.	In the case of global distribution, this algorithm fails.
[23, 26, 31, 79–81]	Histogram thresholding	The quality of segmentation depends on the threshold value.	Deciding the threshold value is a crucial task.
[25, 53, 71, 75]	K-means clustering	Unsupervised segmentation technique used to obtain the same feature regions.	The value of the cluster, i.e., K, must be defined.
[28, 29, 89]	Watershed algorithm	Used for continuous boundary regions extraction. Gives good results on overlapping cells.	The calculation of gradients is complex.
[20, 38, 51, 59, 66, 88]	Marker-controlled watershed	Used to separate overlapped cells.	Does not work on extremely overlapped cells.
[23, 33, 43, 62, 90]	Morphological operation	Mathematical operations are used to separate RBC based on size, texture, boundaries, gradient, circular shape, etc.	High time complexity.
[28, 32, 91–93]	Edge detection algorithm	Excellent results on high contrast and sharp edge blood smear images.	It is a time-consuming process if there are many edges.
[94, 95]	Rule-based segmentation	Required understanding of color, shape, and size of RBC.	RBC's color, size, and shape understanding are required.
[96–98]	Fuzzy rule-based segmentation	Rules need to be designed for segmentation, which is a complex task.	Designing rules is a complex task.
[21, 77, 99, 100]	Hough transform	Used to segment accurate radius and shape of cells.	Computationally expensive in case of a large number of parameters.

computational complexity of the other processes and provides accurate and reliable recognition to unknown, unrecognized data.

To develop a good classification model, a good feature selection method plays a very important role. Classification model processing time and results of classification model depend on selection and type of the number of selected features or attributes. In the literature, several researchers have developed and used different feature selection methods.

To extract the features of haralick textures, mean, entropy, roughness, homogeneity, and standard deviation, Das et al. suggested gray-level co-occurrence matrix [66]. To extract the intensity-based features, Chayadevi and Raju used a color channel intensity algorithm [96]. Rajaraman et al. have given a pretrained model for the feature extraction and the detection of malaria parasites [101]. In this, a pretrained convolutional neural network including AlexNet, VGG-16, Xception, ResNet-50, and DenseNet-121 are used for extracting features from infected and uninfected 27558 cell images. The developed model for feature extraction and malaria parasite detection took more than 24 hours for training and produced 95.9% accuracy for malaria parasite detection in thin blood smear images. To identify the texture features from a blood smear image to detect malaria parasites, Chavan and Sutkar used a histogram-based feature extraction method [102]. The color histogram feature extraction technique is used by [43] for identifying infected erythrocytes from blood smear images. Reference [103] extracted features of RBC size and shape, RBC texture, and parasite shape from the thin blood smear images, and used these features to classify malaria parasite species. For extracting the features from digital microscopic images based on morphological, [43] used a granulometry algorithm.

Various features of extraction and selection techniques implemented by various researchers for malaria blood smear images are shown in Table 6. As evident from Table 6, it is found that researchers used different feature extraction techniques based on their goals. Mostly used feature extraction techniques were color features and texture features. However, some authors recommended morphological feature technique for features extraction from malaria blood smear images [51, 108].

3.5. Parasite Identification and Labelling (Classification). Classification is a technique to identify a pattern that belongs to which class. In this literature, different authors developed different classification techniques to identify a patient, whether he or she is malaria-affected or not. So, there are two classes to detect whether the patient is affected by malaria or not.

Vijayalakshmi and Kanna have introduced a deep learning approach to classify infected and noninfected falciparum malaria. The presented technique was achieved by the visual geometry group (VGG) network and SVM. In this, 1530 malaria digital corpus images have been used for training and testing the model. In this, the transfer learning approach to train the model is used in which we trained the top layer of the model and freeze the rest out of the layers approach applied. For the classification of infected or noninfected falciparum malaria, the given model obtained 93.13% accuracy [8].

For the classification of malaria-infected stages from thin blood smear images, Das et al. used five different classifiers to classify the malaria-infected stages. These five classifiers are Naive Bayes, Logistic regression, Radial Basis Functions (RBF) neural network, Multilayer perceptron neural

TABLE 6: Different techniques used for the extraction of features and selection from malaria blood smear images.

References	Color features				Texture feature							Morphological feature						
	RGB	HSV	YCbCr	Lab	Intensity	CCM	Haralick	GLRLM	GLCM	LBP	Fractal	WT	GT	Entropy	SIFT	Shape	Moments	Area
[24]	✓																	
[104]	✓															✓		
[105]					✓											✓		
[106]		✓					✓		✓									
[96]		✓																
[70]	✓			✓														
[107]	✓															✓		
[108]											✓		✓			✓		
[109]														✓		✓		
[51]								✓						✓		✓		
[100]								✓		✓				✓		✓		
[25]																		
[110]																		
[111]						✓				✓					✓			
[43]																		✓
[38]	✓																	✓
[66]									✓									
[102]																		
[65]									✓									
[21]																		
[20]	✓													✓				✓
[31]	✓																	

RGB—red green blue; HSV—hue, saturation, and value; CCM—color co-occurrence matrix; GLRLM—Gray-level run length matrices; GLCM—gray-level co-occurrence matrix; LBP—local binary pattern; WT—wavelet transform; GT—gradient texture.

network, and classification and regression tree. In this, 888 erythrocytes infected and noninfected image dataset is used. Out of this, 592 labeled images are used to train the classifiers and the remaining are used for testing the classifiers. The experimental results show that among all five classifiers, the multilayer perceptron network has provided better results than the other four classifiers on the 750 images dataset [66].

Seman et al. developed a multilayer perceptron network (MLP) to classify different malaria parasite species from thin blood smear images. This work classified three different species from malaria parasites: *P. falciparum*, *P. malariae*, and *P. vivax* [103]. The authors used the backpropagation algorithm of the MLP network for training and compared the results of the MLP network with Levenberg–Marquardt and Bayesian rule algorithms. MLP network has produced better results as compared to the other two algorithms.

Otsu thresholding method is used by Malihi et al. for the classification of four species of malaria parasites in blood smear images. This technique has provided better results in comparison with other techniques. In this, 363 blood smear images are used and obtained 91% accuracy [43].

Further, Anggraini et al. have classified the different stages of malaria parasites using a Bayesian classifier on 110 thin blood smear images and obtained 93.3% accuracy [112]. Minimum distance classifier technique has been given by Ghate et al. for detecting the presence of malaria parasites using 80 blood smear images and got 83.75% accuracy [39]. Savkare and Narote presented Otsu thresholding, watershed transform, and SVM binary classifier to classify normal and parasite-infected cells [113]. Das et al. have presented the Bayesian approach for automated screening of malaria parasite from microscopic images [51].

Kareem et al. have developed an automated technique for detecting malaria parasites in thin blood smear images. In this, a dataset of more than 200 images is used. Two methods of classification for parasites are used. The first one is based on relative size and morphology, and the second is based on intensity variation. The final results of the developed model have shown an accuracy rate of 87% [36].

Prasad et al. have presented a decision support system approach to classify the infected and noninfected malaria parasites in thin blood smear images. In this, 200 thin blood smear images have been used, and 96% accuracy has been obtained [114]. Rosado et al. have developed a supervised classification technique to detect malaria parasites in blood smears. In this, machine learning (ML) classification 10-fold cross-validation for WBC (white blood cell) and *P. falciparum* trophozoites detection has been performed and got 91.8% accuracy [85].

Mohammed and Abdelrahman have given a technique for detecting and classifying malaria from 160 thin blood smear images taken from the Centre for Disease Control and Prevention (CDC). To extract the RBC from blood smear images, researchers used morphological processing. This technique found the parasites and overlapped cells in the image. Based on the number of RBCs in each image, RBC is classified into two classes: infected and noninfected cells. After that, a normalized cross-correlation algorithm was employed to classify the affected blood smear parasite into

four different malaria species. The given method has produced 95% accuracy for detection [23]. Saiprasath et al. evaluated seven different machine learning algorithms on the same malaria image dataset and concluded that Random Forest outperforms every other algorithm, closely followed by the Ada Boost algorithm [115].

Bibin et al. have given an automated technique to detect malaria parasites in peripheral blood smear images. The given binary classifier is based on deep learning, which used 1978 images to train and test the technique and achieved 96.21% accuracy [106]. Simon et al. suggested a CNN-RNN model for malaria detection. Compared with the CNN-LSTM and CNN-GRU models, the proposed model generated the best results [116]. Dev et al. suggested a hierarchical convolutional network and produced better results than prior studies [117].

Dave et al. used adaptive thresholding, erosion, and dilation operations to diagnose malaria from 117 blood smear microscopic digital images and got 89.88% accuracy [34]. Oliveira et al. have suggested the face detection algorithm to identify Plasmodium parasites from blood samples. In this, a dataset of 1332 blood sample images has been taken and shown 91% accuracy [118].

Mohanty et al. have presented the autoencoder (AE) neural network unsupervised technique to identify malaria in blood smear images. In this, the AE technique has been compared with the SOM technique. The AE technique obtained 87.5% accuracy compared to the 79% accuracy of the SOM technique. 1182 blood smear images have been used to perform experiments [119]. Morales-Lopez et al. suggested the SVM technique for classification problems [120]. Table 7 has listed different types of classification techniques used for the identification of malaria parasites.

After the analysis of Table 7 and literature of malaria parasite classification techniques, it is found that various classification techniques that researchers commonly implement are CNN, SVM, and TL-VGG classifiers.

4. Discussion

In the last decade, a lot of experiments have been done in the area of automated detection of malaria to reach the current state-of-the-art. In this study, different computational methods implemented on various stages of computer-assisted techniques for detecting malaria parasites using blood smear images have been examined. Image acquisition is the first and very important step for automatic detection of the malaria parasite. The present study shows various techniques for acquiring digital blood smear images, but the light microscopy technique is the most widely used and liked technique by researchers. There is a number of computational methods out of which preprocessing is the first step in image analysis.

Preprocessing is one of the crucial stages implemented on acquired digital blood smear images. It plays a crucial role in detecting infected RBC by removing the unwanted noise and producing high contrast digital blood smear images without demolishing the image features. As per the current study, median/mean filter, morphological filter, Laplacian

TABLE 7: Different classification techniques used for the identification of malaria parasites.

Reference	Technique used	Dataset	Accuracy (%)	Limitations/challenges
[103]	Multilayer perceptron network for classification of malaria species.	562 malaria images	89.90	Computation cost is very high.
[113]	Otsu thresholding, watershed transform, and SVM binary classifier for classification of normal and parasite-infected cells.	15 malaria images	93.12	Species detection of malaria is not done. Not suitable for large datasets.
[121]	Comprehensive CAD techniques with 10-fold cross-validation.	1182 malaria images	89.10	Training and testing time is very high for large datasets.
[93]	Suggested SVM technique to find the different stages of infected malaria parasite	530 malaria images	86	Feature scaling is required.
[73]	Used RGB color space model and Otsu algorithm for RBC and parasite segmentation from thin blood smear images.	20 malaria images	92	The unpredictability and imperfections in microscope pictures make precise detection difficult.
[114]	The decision support system for the classification of an infected and noninfected parasite of malaria.	200 malaria images	96	FP rate is 20% and used only thin blood smear images.
[122]	Used minimum distance classifier to detect malaria parasites in blood smear images.	80 malaria images	83.75	Dataset size is very small.
[43]	Used SVM, NM, KNN, 1-NN, and Fisher classifiers to classify different malaria species.	363 malaria images	91	Using a hybrid approach, results can be improved.
[51]	Used Bayesian algorithm for detection of the malaria parasite.	888 malaria images	84	Detect only 1 stage of malaria.
[123]	An artificial neural network has been used to identify the different malaria species from malaria parasites' blood smear images.	200 malaria images	79.7	Performance can be improved by extracting more features.
[96]	Used the neural network method to identify infected RBC from blood smear images.	476 malaria images	94.45	Results can be improved by training the model on a large dataset.
[75]	K-means clustering has been used for the segmentation of malaria parasites cells.	118 malaria images	95	Other types of parasites are not detectable with this technique.
[74]	For the segmentation of RBC, the K-means clustering technique and global threshold technique have been used.	78 malaria images	95.5	Dataset size is minimal.
[66]	For the classification of gametocyte stage and ring stage of malaria species, multilayer perceptron network and 4 other classifiers have been used.	750 malaria images	96.73	By increasing training size, more accurate results can be achieved.
[124]	Used artificial neural network (ANN) for the detection of malaria parasite using morphological features.	7 malaria images	73.57	Achieve better results by increasing dataset size and using 2 or more classifiers.
[85]	Used SVM classifier for WBC and <i>P. falciparum</i> trophozoites detection.	1843 malaria images	91.8	Implemented only with the mobile-based framework.
[125]	Used image processing and artificial intelligence techniques and face detection algorithm to identify plasmodium parasites from blood samples.	1332 malaria images	91	Detected only 1 malaria parasite, and more algorithms can explore to achieve better accuracy.
[106]	Malaria parasite detection using a deep belief network.	1978 malaria images	96.21	The technique was not implemented on a dataset acquired from a mobile phone.
[119]	Used autoencoder neural network technique to identify malaria in blood smear images.	1182 malaria images	87.5	The segmentation technique can be improved.
[101]	Used 6 pretrained CNN for feature extraction and subsequent training for malaria parasite detection in thin blood smear images. This model took more than 24 hours for training.	27558 malaria images	95.9	The model took more than 24 hours for training.
[8]	Used transfer learning approach based on VGG-SVM model to classify infected and noninfected falciparum malaria parasite.	1530 malaria images	93.13	A trained model can recognize only 1 falciparum malaria parasite.
[126]	Used CNN based deep learning model (VGGNet-16 architecture) for malaria parasite detection.	27558 malaria images	95.03	Results can be improved by implementing the VGG-19 architecture.

TABLE 7: Continued.

Reference	Technique used	Dataset	Accuracy (%)	Limitations/challenges
[127]	Used custom CNN that consists of three fully connected convolutional layers.	17460 malaria images	95	A model can test on more computing power systems for better results.

filter, Susan filter, and Gaussian low-pass filter are mostly used techniques by researchers to remove unwanted noise and increase the contrast of the images. Segmentation is the next stage after the preprocessing, which is used to segment the RBC to detect malaria parasites using blood smear images to facilitate the classification process. As per the study of literature, mostly used segmentation techniques by researchers are as follows: (i) Otsu thresholding for segmentation of parasite RBC; (ii) Marker-controlled watershed and Edge detection algorithm is used at the segmentation phase; and (iii) for the segmentation of overlapping cells, the watershed algorithm has been widely used.

After segmentation, blood smear images have been classified to diagnose malaria-infected or not infected by feature extraction and selection techniques. As per the study, Color features, Texture features, and Morphological features have been mostly used feature extraction techniques for early diagnosis of malaria from blood smear images. From the literature, it has been observed that the maximum accuracy of 95.03% has been achieved by CNN based deep learning model in comparison with the VGG-SVM model [8, 101, 126].

A thorough review of the literature on automated analysis of malaria parasite using blood smear images yielded the following challenges and future directions:

The accuracy of an automatic image classification model depends upon multiple aspects such as analysis of the digital blood smear image depend on the staining method, magnification factor of an image, condition of nearest environment where the digital image has been collected like the background of the image, light in the room, and most important quality and position of the camera. Therefore, a standard digital blood smear dataset is necessary to test and validate the model to obtain efficient and reliable results.

Many researchers have performed their experiments and published their articles in the same area. Moreover, an automated computational-based computer vision method, which should be efficient and effective for automated detection of the malaria parasite from blood smear images, needs to be improvised according to the requirement of the community.

The community requires (i) standard image dataset because researchers' datasets are mostly unstandardized. The digital blood smear dataset depends on the characteristics and quality of the microscope as all digital images of blood smears are taken by a digital camera attached to a microscope. So, a well-standardized dataset is most important for a machine learning algorithm for automated detection of malaria. (ii) In the literature, developed methods can recognize only one type of malaria parasite [8]. But, the patient may be affected by more than one

parasites species. Hence, there is a need for such a model that can recognize different types of malaria parasites. (iii) To classify malaria parasites from blood smear images, authors trained the machines with different models and techniques. The training model is taking a long time to learn [66]. Hence, there is a necessity to reduce the training time to train the classification models. (iv) In literature, developed models by different researchers analyze the blood smear digital images that are taken from a camera that is attached with a microscope [4, 8]. Hence, there is a demand for a model to analyze thin blood smears images acquired exclusively with smart phones [43]. (v) A technique that different authors use to diagnose malaria is invasive, in which an injection syringe takes a blood sample. Therefore, there is a requirement for a noninvasive technique that can be used to detect malaria [128]. (vi) After the analysis of Table 7, it has been found that all the existing state-of-the-art techniques used to detect malaria from microscopic blood smear images are not very accurate. Each technique has some limitations and challenges. Therefore, there is a necessity for an automatic technique that can improve the accuracy for the detection of malaria parasites and it will also help in early detection of malaria and reduce the mortality rate in future.

5. Conclusion

This study is a solid starting point for researchers who want to look into automated blood smear analysis to detect malaria. This study reviews and discusses computer vision and image analysis works that target the automated detection of malaria on blood smear images. In this paper, we have discussed the present facts of necessary components of computer-assisted technique: (i) acquisition of image dataset, (ii) preprocessing, (iii) segmentation of RBC, (iv) feature extraction and selection, and (v) classification, which have been used to diagnose malaria parasite from blood smear images suggested by various researchers. Digital blood smear images taken from a microscope may affect how and which malaria parasites are detected. After analyzing segmentation and classification state of the art techniques, it has been observed that future computer-assisted techniques should be based on standard datasets and magnification factors to detect malaria parasites. The complexity of different classifiers of machine learning that are based on deep learning increases as the number of layers increases. To achieve efficient and reliable results, a large dataset is required for training and testing. With the help of computational methods such as data augmentation and deep learning methods, the computer-assisted method can obtain better results.

However, some state-of-the-art techniques are presented in the literature, but still, there is a huge scope of future work, which may help the microbiologists in the detection and diagnosis of the malaria parasite at an earlier stage to reduce the mortality rate such as (i) different computational methods that are used to collect blood smear images physically can be studied more to enhance the segmentation results to detect infected RBC very effectively. Hence, an efficient computational method of infected RBC segmentation can be developed. (ii) Various feature extraction methods such as color features, texture features, and morphological features [51, 106, 108] can be analyzed more, which will be very helpful for the development of an efficient automated computer-assisted system to detect infected malaria RBC using blood smear images. (iii) To classify malaria blood smear images, mostly implemented techniques are SVM, K-means, and VGG classifiers. Still, there is a vast scope to implement customize CNN algorithms to detect infected malaria RBC with high accuracy. If the CNN model is implemented on blood smear images at a minimum magnification factor for classification, it may decrease the cost and time complexity of the system.

In the field of malaria detection from blood smear images, the contribution of many research publications is noteworthy. However, this study has tried to present opinions to the microbiologists and technical community. It will be very helpful for them to generate an effective and efficient computer-assisted technique for malaria detection at an early stage. [129, 130].

Data Availability

Dataset is available at <https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria>.

Conflicts of Interest

The authors stated that there are no conflicts of interest.

Acknowledgments

The authors would like to acknowledge Taif University Researchers Supporting Project Number (TURSP-2020/125), Taif University, Taif, Saudi Arabia, http://shodh.inflibnet.ac.in:8080/jspui/bitstream/123456789/8785/1/shankar_phdeng17050_synopsis.pdf.

References

- [1] Who, "World health organization report on malaria," Report/2021, World health organization, Geneva, Switzerland, 2021.
- [2] Incidence of Malaria in India, "Incidence of malaria in India," 2020, <https://www.malariasite.com/malaria-india>.
- [3] F. E. Cox, "History of the discovery of the malaria parasites and their vectors," *Parasites & Vectors*, vol. 3, no. 1, pp. 5–9, 2010.
- [4] Z. May and S. S. A. M. Aziz, "Automated quantification and classification of malaria parasites in thin blood smears," in *Proceedings of the 2013 IEEE International Conference on Signal and Image Processing Applications*, pp. 369–373, Melaka, Malaysia, October 2013.
- [5] Who, "Transmission of malaria," 2020, <https://www.who.int/features/qa/10/en>.
- [6] H. M. Gilles, "Management of severe and complicated malaria," *A Practical Handbook*, World Health Organization, Geneva, Switzerland, 1991.
- [7] S. Murphy and J. Breman, "Gaps in the childhood malaria burden in Africa: cerebral malaria, neurological sequelae, anemia, respiratory distress, hypoglycemia, and complications of pregnancy," *The American Journal of Tropical Medicine and Hygiene*, vol. 64, no. 1_suppl, pp. 57–67, 2001.
- [8] A. Vijayalakshmi and B. R. Kanna, "Deep learning approach to detect malaria from microscopic images," *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 15297–15317, 2020.
- [9] A. Loddo, C. Di Ruberto, and M. Kocher, "Recent advances of malaria parasites detection systems based on mathematical morphology," *Sensors*, vol. 18, no. 2, p. 513, 2018.
- [10] N. Tangpukdee, C. Duangdee, P. Wilairatana, and S. Krudsood, "Malaria diagnosis: a brief review," *Korean Journal of Parasitology*, vol. 47, no. 2, p. 93, 2009.
- [11] Ncbi, "Malaria diagnosis: a brief review," 2020, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2688806>.
- [12] A. Rosebrock, "Microscopic thick and thin blood smears examination," 2020, <http://medicostenerife.es/news/inteligencia-artificial-deep-learning-and-medical-image-analysis-with-keras-for-malaria-deteccion-97-effectiveness>.
- [13] I. Bates, V. Bekoe, and A. Asamoah-Adu, "Improving the accuracy of malaria-related laboratory tests in Ghana," *Malaria Journal*, vol. 3, no. 1, pp. 1–5, 2004.
- [14] A. Tankeshwar, "Rapid diagnosis test," 2019, <https://microbeonline.com/rdts-malaria-diagnosis-principle-results-advantages>.
- [15] V. N. Orish, V. F. De-Gaulle, and A. O. Sanyaolu, "Interpreting rapid diagnostic test (RDT) for Plasmodium falciparum," *BMC Research Notes*, vol. 11, no. 1, pp. 850–856, 2018.
- [16] Z. Jan, A. Khan, M. Sajjad, K. Muhammad, S. Rho, and I. Mehmood, "A review on automated diagnosis of malaria parasite in microscopic blood smears images," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 9801–9826, 2018.
- [17] K. M. F. Fuhad, J. F. Tuba, M. R. A. Sarker, S. Momen, N. Mohammed, and T. Rahman, "Deep learning based automatic malaria parasite detection from blood smear and its smartphone based application," *Diagnostics*, vol. 10, no. 5, p. 329, 2020.
- [18] C. Wongsrichanalai, F. Kawamoto, M. Hommel, and P. G. Kremsner, "Fluorescent microscopy and fluorescent labelling for malaria diagnosis," *Encycl. Malar*, pp. 1–7, Springer, New York, NY, USA, 2021.
- [19] M. Umer, S. Sadiq, M. Ahmad, S. Ullah, G. S. Choi, and A. Mehmood, "A novel stacked cnn for malarial parasite detection in thin blood smear images," *IEEE Access*, vol. 8, pp. 93782–93792, 2020.
- [20] S. S. Devi, A. Roy, J. Singha, S. A. Sheikh, and R. H. Laskar, "Malaria infected erythrocyte classification based on a hybrid classifier using microscopic images of thin blood smear," *Multimedia Tools and Applications*, vol. 77, no. 1, pp. 631–660, 2018.
- [21] Y. Dong, Z. Jiang, H. Shen, W. David Pan, L. A. Williams, and V. V. B. Redd, "Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells," in *Proceedings of the 2017 IEEE EMBS International*

- Conference on Biomedical & Health Informatics (BHI)*, pp. 101–104, Orlando, FL, USA, February 2017.
- [22] S. M. Parsel, S. A. Gustafson, E. Friedlander et al., “Malaria over-diagnosis in Cameroon: diagnostic accuracy of Fluorescence and Staining Technologies (FAST) Malaria Stain and LED microscopy versus Giemsa and bright field microscopy validated by polymerase chain reaction,” *Infectious diseases of poverty*, vol. 6, no. 1, pp. 32–39, 2017.
- [23] H. A. Mohammed and I. A. M. Abdelrahman, “Detection and classification of malaria in thin blood slide images,” in *Proceedings of the 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)*, pp. 1–5, Khartoum, Sudan, January 2017.
- [24] D. Yang, G. Subramanian, J. Duan et al., “A portable image-based cytometer for rapid malaria detection and quantification,” *PLoS One*, vol. 12, no. 6, Article ID e0179161, 2017.
- [25] A. Nanoti, S. Jain, C. Gupta, and G. Vyas, “Detection of malaria parasite species and life cycle stages using microscopic images of thin blood smear,” in *Proceedings of the 2016 International Conference on Inventive Computation Technologies (ICICT)*, vol. 1, pp. 1–6, Coimbatore, India, October 2016.
- [26] Z. Liang, A. Powell, I. Ersoy, M. Poostchi, K. Silamut, and K. Palani, “CNN-based image analysis for malaria diagnosis,” in *Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 493–496, Shenzhen, China, December 2016.
- [27] K. T. Fn, T. Daniel, E. Pierre, T. Emmanuel, and B. Philippe, “Automated diagnosis of malaria in tropical areas using 40X microscopic images of blood smears,” *International Journal of Biometric and Bioinformatics*, vol. 10, no. 2, p. 12, 2016.
- [28] J.-D. Kim, K.-M. Nam, C.-Y. Park, Y.-S. Kim, and H.-J. Song, “Automatic detection of malaria parasite in blood images using two parameters,” *Technology and Health Care*, vol. 24, no. s1, pp. S33–S39, 2016.
- [29] S. S. Savkare and S. P. Narote, “Automated system for malaria parasite identification,” in *Proceedings of the 2015 international conference on communication, information & computing technology (ICCICT)*, pp. 1–4, Mumbai, India, January 2015.
- [30] C. W. Pirnstill and G. L. Coté, “Malaria diagnosis using a mobile phone polarized microscope,” *Scientific Reports*, vol. 5, no. 1, pp. 1–13, 2015.
- [31] D. L. Omucheni, K. A. Kaduki, W. D. Bulimo, and H. K. Angeyo, “Application of principal component analysis to multispectral-multimodal optical image analysis for malaria diagnostics,” *Malaria Journal*, vol. 13, no. 1, pp. 1–11, 2014.
- [32] B. Maiseli, J. Mei, H. Gao, and S. Yin, “An automatic and cost-effective parasitemia identification framework for low-end microscopy imaging devices,” in *Proceedings of the 2014 International Conference on Mechatronics and Control (ICMC)*, pp. 2048–2053, Jinzhou, China, July 2014.
- [33] M. C. Mushabe, R. Dendere, and T. S. Douglas, “Automated detection of malaria in Giemsa-stained thin blood smears,” in *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3698–3701, Osaka, Japan, July 2013.
- [34] S. Moon, S. Lee, H. Kim et al., “An image analysis algorithm for malaria parasite stage classification and viability quantification,” *PLoS One*, vol. 8, no. 4, Article ID e61812, 2013.
- [35] S. Bhowmick, D. K. Das, A. K. Maiti, and C. Chakraborty, “Structural and textural classification of erythrocytes in anaemic cases: a scanning electron microscopic study,” *Micron*, vol. 44, pp. 384–394, 2013.
- [36] S. Kareem, I. Kale, and R. C. S. Morling, “Automated malaria parasite detection in thin blood films: A hybrid illumination and color constancy insensitive, morphological approach,” in *Proceedings of the 2012 IEEE Asia Pacific Conference on Circuits and Systems*, pp. 240–243, Kaohsiung, Taiwan, December 2012.
- [37] S. Dabo-Niang and J. T. Zoueu, “Combining kriging, multispectral and multimodal microscopy to resolve malaria-infected erythrocyte contents,” *Journal of Microscopy*, vol. 247, no. 3, pp. 240–251, 2012.
- [38] N. Ahirwar, S. Pattnaik, and B. Acharya, “Advanced image analysis based system for automatic detection and classification of malarial parasite in blood images,” *International Journal of Information Technology and Knowledge Management*, vol. 5, no. 1, pp. 59–64, 2012.
- [39] S. Mavandadi, S. Dimitrov, S. Feng et al., “Distributed medical image analysis and diagnosis through crowd-sourced games: a malaria case study,” *PLoS One*, vol. 7, no. 5, Article ID e37245, 2012.
- [40] A. Simon, R. Vinayakumar, V. Sowmya, and K. P. Soman, “Shallow cnn with lstm layer for tuberculosis detection in microscopic images,” *Machine learning for Biomedical Applications*, 2019.
- [41] J. Soni, N. Mishra, and C. Kamargaonkar, “Automatic differentiation between RBC and malarial parasites based ON morphology with first order features using image processing,” *International Journal of Advances in Engineering & Technology*, vol. 1, no. 5, p. 290, 2011.
- [42] D. N. Breslauer, R. N. Maamari, N. A. Switz, W. A. Lam, and D. A. Fletcher, “Mobile phone based clinical microscopy for global health applications,” *PLoS One*, vol. 4, no. 7, Article ID e6320, 2009.
- [43] L. Malihi, K. Ansari-Asl, and A. Behbahani, “Malaria parasite detection in giemsa-stained blood cell images,” in *Proceedings of the 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*, pp. 360–365, Zanjan, Iran, September 2013.
- [44] A. Mehrjou, T. Abbasian, and M. Izadi, “Automatic malaria diagnosis system,” in *Proceedings of the 2013 1st RSI/ISM International Conference on Robotics and Mechatronics (ICRoM)*, pp. 205–211, Tehran, Iran, February 2013.
- [45] R. Rosnelly, “Identification of malaria disease and its stadium based on digital image processing,” 2016.
- [46] M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger, and G. Thoma, “Image analysis and machine learning for detecting malaria,” *Translational Research*, vol. 194, pp. 36–55, 2018.
- [47] J. Gatc, F. Maspiyanti, D. Sarwinda, and A. M. Arymurthy, “Plasmodium parasite detection on red blood cell image for the diagnosis of malaria using double thresholding,” in *Proceedings of the 2013 international conference on advanced computer science and information systems (ICACSIS)*, pp. 381–385, Sanur Bali, Indonesia, September 2013.
- [48] P. Rakshit and K. Bhowmik, “Detection of presence of parasites in human RBC in case of diagnosing malaria using image processing,” in *Proceedings of the 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013)*, pp. 329–334, Shimla, India, December 2013.
- [49] M. I. Khan, B. Acharya, B. K. Singh, and J. Soni, “Content based image retrieval approaches for detection of malarial parasite in blood images,” *International Journal of Biometric and Bioinformatics*, vol. 5, no. 2, p. 97, 2011.

- [50] N. E. Ross, C. J. Pritchard, D. M. Rubin, and A. G. Dusé, "Automated image processing method for the diagnosis and classification of malaria on thin blood smears," *Medical, & Biological Engineering & Computing*, vol. 44, no. 5, pp. 427–436, 2006.
- [51] D. K. Das, M. Ghosh, M. Pal, A. K. Maiti, and C. Chakraborty, "Machine learning approach for automated screening of malaria parasite using light microscopic images," *Micron*, vol. 45, pp. 97–106, 2013.
- [52] S. Kaewkamnerd, C. Uthapibull, A. Intarapanich, M. Pannarut, S. Chaotheing, and S. Tongshima, "An automatic device for detection and classification of malaria parasite species in thick blood film," *BMC Bioinformatics*, vol. 13, no. 17, pp. 1–10, 2012.
- [53] A.-N. Aimi Salihah, M. Yusoff, and M. Zeehaida, "Colour image segmentation approach for detection of malaria parasites using various colour models and k-means clustering," *WSEAS Transactions on Biology and Biomedicine*, vol. 10, 2013.
- [54] Y. Purwar, S. L. Shah, G. Clarke, A. Almgairi, and A. Muehlenbachs, "Automated and unsupervised detection of malarial parasites in microscopic images," *Malaria Journal*, vol. 10, no. 1, pp. 1–11, 2011.
- [55] G. Díaz, F. A. González, and E. Romero, "A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images," *Journal of Biomedical Informatics*, vol. 42, no. 2, pp. 296–307, 2009.
- [56] J. Somasekar, B. E. Reddy, E. K. Reddy, and C.-H. Lai, "An image processing approach for accurate determination of parasitemia in peripheral blood smear images," *International Journal of Computers and Applications*, vol. 1, pp. 23–28, 2011.
- [57] M.-H. Tsai, S.-S. Yu, Y.-K. Chan, and C.-C. Jen, "Blood smear image based malaria parasite and infected-erythrocyte detection and segmentation," *Journal of Medical Systems*, vol. 39, no. 10, pp. 1–14, 2015.
- [58] L. Gitonga, D. M. Memeu, K. A. Kaduki, A. C. K. Mjomba, and N. S. Muriuki, "Determination of plasmodium parasite life stages and species in images of thin blood smears using artificial neural networks," *Open Journal of Clinical Diagnostics*, vol. 4, 2014.
- [59] S. Kareem, R. C. S. Morling, and I. Kale, "A novel method to count the red blood cells in thin blood films," in *Proceedings of the 2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, pp. 1021–1024, Rio de Janeiro, Brazil, May 2011.
- [60] K. M. Khatri, V. R. Ratnaparkhe, S. S. Agrawal, and A. S. Bhalchandra, "Image processing approach for malaria parasite identification," *International Journal of Computer Applications® (IJCA)*, 2013.
- [61] F. Sheeba, R. Thamburaj, J. J. Mammen, and A. K. Nagar, "Detection of plasmodium falciparum in peripheral blood smear images," *Advances in Intelligent Systems and Computing*, vol. 202, pp. 289–298, 2013.
- [62] J. E. Arco, J. M. Górriz, J. Ramírez, I. Álvarez, and C. G. Puntonet, "Digital image analysis for automatic enumeration of malaria parasites using morphological operations," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3041–3047, 2015.
- [63] Y.-W. Hung, C.-L. Wang, C.-M. Wang et al., "Parasite and infected-erythrocyte image segmentation in stained blood smears," *Journal of Medical and Biological Engineering*, vol. 35, no. 6, pp. 803–815, 2015.
- [64] S. K. Reni, I. Kale, and R. Morling, "Analysis of thin blood images for automated malaria diagnosis," in *Proceedings of the 2015 E-Health and Bioengineering Conference (EHB)*, pp. 1–4, Iași, Romania, December 2015.
- [65] M. I. Razzak, "Automatic detection and classification of malarial parasite," *International Journal of Biometric and Bioinformatics*, vol. 9, no. 1, pp. 1–12, 2015.
- [66] D. K. Das, A. K. Maiti, and C. Chakraborty, "Automated system for characterization and classification of malaria-infected stages using light microscopic images of thin blood smears," *Journal of Microscopy*, vol. 257, no. 3, pp. 238–252, 2015.
- [67] N. Abbas and D. Mohamad, "And others, "Microscopic RGB color images enhancement for blood cells segmentation in YCBCR color space for k-means clustering," *Journal of Theoretical and Applied Information Technology*, vol. 55, no. 1, pp. 117–125, 2013.
- [68] M. Brückner, K. Becker, J. Popp, and T. Frosch, "Fiber array based hyperspectral Raman imaging for chemical selective analysis of malaria-infected red blood cells," *Analytica Chimica Acta*, vol. 894, pp. 76–84, 2015.
- [69] V. Makkapati and R. M. Rao, "Segmentation of malaria parasites in peripheral blood smear images," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1361–1364, Taipei, Taiwan, April 2009.
- [70] S. Mandal, A. Kumar, J. Chatterjee, M. Manjunatha, and A. K. Ray, "Segmentation of blood smear images using normalized cuts for detection of malarial parasites," in *Proceedings of the 2010 Annual IEEE India Conference (INDICON)*, pp. 1–4, Kolkata, India, December 2010.
- [71] A. S. A. Nasir, M. Y. Mashor, and Z. Mohamed, "Segmentation based approach for detection of malaria parasites using moving k-means clustering," in *Proceedings of the 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences*, pp. 653–658, Langkawi, Malaysia, December 2012.
- [72] S. Bhatia, "New improved technique for initial cluster centers of K means clustering using Genetic Algorithm," in *Proceedings of the International Conference for Convergence for Technology-2014*, pp. 1–4, Pune, India, April 2014.
- [73] V. V. Panchbhai, L. B. Damahe, A. V. Nagpure, and P. N. Chopkar, "RBCs and parasites segmentation from thin smear blood cell images," *International Journal of Image, Graphics and Signal Processing*, vol. 4, no. 10, pp. 54–60, 2012.
- [74] S. S. Savkare and S. P. Narote, "Blood cell segmentation from microscopic blood images," in *Proceedings of the 2015 International Conference on Information Processing (ICIP)*, pp. 502–505, Pune, India, December 2015.
- [75] N. A. Khan, H. Pervaz, A. K. Latif, and A. Musharraf, "Unsupervised identification of malaria parasites using computer vision," in *Proceedings of the 2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 263–267, Chon Buri, Thailand, May 2014.
- [76] V. Acharya, V. Ravi, T. D. Pham, and C. Chakraborty, "Peripheral blood smear analysis using automated computer-aided diagnosis system to identify Acute myeloid leukemia," *IEEE Transactions on Engineering Management*, pp. 1–14, 2021.
- [77] C. Ma, P. Harrison, L. Wang, and R. L. Coppel, "Automated estimation of parasitaemia of Plasmodium yoelii-infected mice by digital image analysis of Giemsa-stained thin blood smears," *Malaria Journal*, vol. 9, no. 1, pp. 348–349, 2010.

- [78] V. V. Makkapati and R. M. Rao, "Ontology-based malaria parasite stage and species identification from peripheral blood smear images," in *Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6138–6141, Boston, MA, USA, August 2011.
- [79] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, Boston, MA, USA, May 2015.
- [80] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *Proceedings of the International MICCAI brainlesion workshop*, pp. 178–190, Quebec, QC, Canada, September 2017.
- [81] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [82] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Munich, Germany, October 2015.
- [83] J. Dai, K. He, and J. Sun, "Instance-aware semantic segmentation via multi-task network cascades," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3150–3158, Las Vegas, NV, USA, June 2016.
- [84] F. Visin, A. Romero, K. Cho, M. Matteucci, M. Ciccone, and K. Kastner, "Reseg: a recurrent neural network-based model for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 41–48, Las Vegas, NV, USA, June 2016.
- [85] L. Rosado, J. M. C. d. Costa, D. Elias, and J. S. Cardoso, "Automated detection of malaria parasites on thick blood smears via mobile devices," *Procedia Computer Science*, vol. 90, pp. 138–144, 2016.
- [86] S. S. Devi, A. Roy, M. Sharma, and R. H. Laskar, "kNN classification based erythrocyte separation in microscopic images of thin blood smear," in *Proceedings of the 2016 2nd International Conference on Computational Intelligence and Networks (CINE)*, pp. 69–7272, Bhubaneswar, India, January 2016.
- [87] H. Lee and Y.-P. P. Chen, "Cell morphology based classification for red cells in blood smear images," *Pattern Recognition Letters*, vol. 49, pp. 155–161, 2014.
- [88] G. P. Gopakumar, M. Swetha, G. Sai Siva, and G. R. K. Sai Subrahmanyam, "Convolutional neural network-based malaria diagnosis from focus stack of blood smear images acquired using custom-built slide scanner," *Journal of Biophotonics*, vol. 11, no. 3, Article ID e201700003, 2018.
- [89] J. M. Sharif, M. F. Miswan, M. A. Ngadi, M. S. H. Salam, and M. M. bin Abdul Jamil, "Red blood cell segmentation using masking and watershed algorithm: a preliminary study," in *Proceedings of the 2012 International Conference on Biomedical Engineering (ICoBE)*, pp. 258–262, Penang, Malaysia, February 2012.
- [90] S. Punitha, P. Logeshwari, P. Sivaranjani, and S. Priyanka, "Detection of malarial parasite in blood using image processing," *Asian J. Appl. Sci. Technol*, vol. 1, no. 2, pp. 211–213, 2017.
- [91] E. Komagal, K. S. Kumar, and A. Vigneswaran, "Recognition and classification of malaria plasmodium diagnosis," *International Journal of Engineering Research and Technology*, vol. 2, no. 1, pp. 1–4, 2013.
- [92] J. Somasekar and B. Esvara Reddy, "Segmentation of erythrocytes infected with malaria parasites for the diagnosis using microscopy imaging," *Computers & Electrical Engineering*, vol. 45, pp. 336–351, 2015.
- [93] S. K. Kumarasamy, S. H. Ong, and K. S. W. Tan, "Robust contour reconstruction of red blood cells and parasites in the automated identification of the stages of malarial infection," *Machine Vision and Applications*, vol. 22, no. 3, pp. 461–469, 2011.
- [94] S. W. S. Sio, W. Sun, S. Kumar et al., "MalariaCount: an image analysis-based program for the accurate determination of parasitemia," *Journal of Microbiological Methods*, vol. 68, no. 1, pp. 11–18, 2007.
- [95] S. S. Savkare and S. P. Narote, "Automatic system for classification of erythrocytes infected with malaria and identification of parasite's life stage," *Procedia Technology*, vol. 6, pp. 405–410, 2012.
- [96] M. Chayadevi and G. Raju, "Usage of art for automatic malaria parasite identification based on fractal features," *International Journal of Video & Image Processing and Network Security*, vol. 4, pp. 7–15, 2014.
- [97] M. L. Chayadevi and G. T. Raju, "Automated colour segmentation of malaria parasite with fuzzy and fractal methods," in *Computational Intelligence in Data Mining-Volume 3*, pp. 53–63, Springer, Salmon, NY, USA, 2015.
- [98] M. Ghosh, D. Das, C. Chakraborty, and A. K. Ray, "Quantitative characterisation of Plasmodium vivax in infected erythrocytes: a textural approach," *International Journal of Artificial Intelligence and Soft Computing*, vol. 3, no. 3, pp. 203–221, 2013.
- [99] Z. Zhang, L. L. Sharon Ong, K. Fang et al., "Image classification of unlabeled malaria parasites in red blood cells," in *Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3981–3984, Orlando, FL, USA, October 2016.
- [100] V. Muralidharan, Y. Dong, and W. D. Pan, "A comparison of feature selection methods for machine learning based automatic malarial cell recognition in wholeslide images," in *Proceedings of the 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 216–219, Las Vegas, NV, USA, February 2016.
- [101] S. Rajaraman, S. K. Antani, M. Poostchi et al., "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," *PeerJ*, vol. 6, Article ID e4568, 2018.
- [102] S. N. Chavan and A. M. Sutkar, "Malaria disease identification and analysis using image processing," *Int. J. Comput. Technol*, vol. 1, no. 6, pp. 218–223, 2014.
- [103] N. A. Seman, N. A. M. Isa, L. C. Li, Z. Mohamed, U. K. Ngah, and K. Z. Zamli, "Classification of malaria parasite species based on thin blood smears using multilayer perceptron network," *International Journal of Computer Integrated Manufacturing*, vol. 16, no. 1, pp. 46–52, 2008.
- [104] I. Suwalka, A. Sanadhya, A. Mathur, and M. S. Chouhan, "Identify malaria parasite using pattern recognition technique," in *Proceedings of the 2012 International Conference on Computing, Communication and Applications*, pp. 1–4, Dindigul, India, February 2012.
- [105] S. Kareem, I. Kale, and R. C. S. Morling, "Automated P. falciparum detection system for post-treatment malaria diagnosis using modified annular ring ratio method," in *Proceedings of the 2012 UKSim 14th International Conference*

- on *Computer Modelling and Simulation*, pp. 432–436, Cambridge, UK, March 2012.
- [106] D. Bibin, M. S. Nair, and P. Punitha, “Malaria parasite detection from peripheral blood smear images using deep belief networks,” *IEEE Access*, vol. 5, pp. 9099–9108, 2017.
- [107] D. K. Das, C. Chakraborty, B. Mitra, A. K. Maiti, and A. K. Ray, “Quantitative microscopy approach for shape-based erythrocytes characterization in anaemia,” *Journal of Microscopy*, vol. 249, no. 2, pp. 136–149, 2013.
- [108] L. Yunda, A. Alarcón, and J. Millán, “Automated image analysis method for p-vivax malaria parasite detection in thick film blood images,” *Sistemas y Telemática*, vol. 10, no. 20, pp. 9–25, 2012.
- [109] A. Ajala, A. Funmilola, F. Fenwa, D. Olusayo, A. Aku, and A. Micheal, “Comparative analysis of different types of malaria diseases using first order features,” *International Journal of Applied Information Systems*, vol. 8, no. 3, pp. 20–26, 2015.
- [110] M. Maity, A. K. Maity, P. K. Dutta, and C. Chakraborty, “A web-accessible framework for automated storage with compression and textural classification of malaria parasite images,” *International Journal of Computer Application*, vol. 52, no. 15, pp. 31–39, 2012.
- [111] N. Linder, R. Turkki, M. Walliander et al., “A malaria diagnostic tool based on computer vision screening and visualization of Plasmodium falciparum candidate areas in digitized blood smears,” *PLoS One*, vol. 9, no. 8, Article ID e104855, 2014.
- [112] D. Anggraini, A. S. Nugroho, C. Pratama, I. E. Rozi, V. Pragesjvara, and M. Gunawan, “Automated status identification of microscopic images obtained from malaria thin blood smears using Bayes decision: a study case in Plasmodium falciparum,” in *Proceedings of the 2011 International Conference on Advanced Computer Science and Information Systems*, pp. 347–352, Jakarta, Indonesia, December 2011.
- [113] S. S. Savkare and S. P. Narote, “Automatic detection of malaria parasites for estimating parasitemia,” *International Journal of Computer Science and Security*, vol. 5, no. 3, p. 310, 2011.
- [114] K. Prasad, J. Winter, U. M. Bhat, R. V. Acharya, and G. K. Prabhu, “Image analysis approach for development of a decision support system for detection of malaria parasites in thin blood smear images,” *Journal of Digital Imaging*, vol. 25, no. 4, pp. 542–549, 2012.
- [115] G. B. Saiprasath, R. Naren Babu, J. ArunPriyan, R. Vinayakumar, V. Sowmya, and K. P. Soman, “Performance comparison of machine learning algorithms for malaria detection using microscopic images,” *IJRAR19RP014 Int. J. Res. Anal. Rev.(IJRAR)*, vol. 6, no. 1, 2019.
- [116] A. Simon, R. Vinayakumar, V. Sowmya, K. P. Soman, and E. A. A. Gopalakrishnan, “A deep learning approach for patch-based disease diagnosis from microscopic images,” in *Classification Techniques for Medical Image Analysis and Computer Aided Diagnosis*, pp. 109–127, Elsevier, Amsterdam, The Netherlands, 2019.
- [117] K. Dev, S. A. Khowaja, A. S. Bist, V. Saini, and S. Bhatia, “Triage of potential covid-19 patients from chest x-ray images using hierarchical convolutional networks,” *Neural Computing & Applications*, no. –16, p. 1, 2021.
- [118] I. R. Dave and K. P. Upla, “Computer aided diagnosis of malaria disease for thin and thick blood smear microscopic images,” in *Proceedings of the 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 561–565, Noida, India, February 2017.
- [119] I. Mohanty, P. A. Pattanaik, and T. Swarnkar, “Automatic detection of malaria parasites using unsupervised techniques,” in *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering*, pp. 41–49, Palladam, India, May 2018.
- [120] H. Morales-Lopez, I. Cruz-Vega, and J. Rangel-Magdaleno, “Cataract detection and classification systems using computational intelligence: a survey,” *Archives of Computational Methods in Engineering*, vol. 28, pp. 1–14, 2020.
- [121] P. A. Pattanaik, M. Mittal, and M. Z. Khan, “Unsupervised deep learning cad scheme for the detection of malaria in blood smear microscopic images,” *IEEE Access*, vol. 8, pp. 94936–94946, 2020.
- [122] D. A. Ghate, C. Jadhav, and N. U. Rani, “Automatic detection of malaria parasite from blood images,” *International Journal of Computer Science and Applications*, vol. 1, 2012.
- [123] D. M. Memeu, *A Rapid Malaria Diagnostic Method Based on Automatic Detection and Classification of Plasmodium Parasites in Stained Thin Blood Smear Images*, University of Nairobi, Nairobi, Kenya, 2014.
- [124] S. T. Khot and R. K. Prasad, “Optimal computer based analysis for detecting malarial parasites,” in *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA)*, pp. 69–80, Durgapur, India, June 2015.
- [125] A. D. Oliveira, C. Prats, M. Espasa, F. Zarzuela Serrat, and C. Montañola Sales, “The malaria system microApp: a new, mobile device-based tool for malaria diagnosis,” *JMIR Res. Protoc*, vol. 6, no. 4, Article ID e6758, 2017.
- [126] A. Sharma, C. Vaishampayan, K. Santlani, M. Sunhare, M. Arya, and S. Gupta, “Malaria parasite detection using deep learning,” *International Journal for Research in Applied Science and Engineering Technology*, vol. 8, no. 5, pp. 163–168, 2020.
- [127] D. Shah, K. Kawale, M. Shah, S. Randive, and R. Mapari, “Malaria parasite detection using deep learning:(beneficial to humankind),” in *Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 984–988, Madurai, India, May 2020.
- [128] M. Osman, H. Salih, O. Salih, N. Abdalaah, and M. Khider, “Design and implementation of non-invasive malaria detection system,” in *Proceedings of the 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, pp. 1–4, Khartoum, Sudan, August 2018.
- [129] L. Zou, J. Chen, J. Zhang, and N. Garcia, “Malaria cell counting diagnosis within large field of view,” in *Proceedings of the 2010 International Conference on Digital Image Computing: Techniques and Applications*, pp. 172–177, Sydney, NSW, Australia, December 2010.
- [130] K. Adi, S. Pujiyanto, R. Gernowo, A. Pamungkas, and A. B. Putranto, “Identifying the developmental phase of plasmodium falciparum in malaria-infected red blood cells using adaptive color segmentation and back propagation neural network,” *International Journal of Applied Engineering Research*, vol. 11, pp. 8754–8759, 2016.

Research Article

Research on Effect of Load Stimulation Change on Heart Rate Variability of Women Volleyball Athletes

Ludi Liao and Jianying Li 

School of Physical Education Shanxi University, Taiyuan, Shanxi 030006, China

Correspondence should be addressed to Jianying Li; lji195912@163.com

Received 10 January 2022; Revised 21 January 2022; Accepted 26 January 2022; Published 19 March 2022

Academic Editor: Jie Liu

Copyright © 2022 Ludi Liao and Jianying Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. To explore the effect of different training load stimulation on heart rate variability level of Chinese elite female volleyball players. Through two-year follow-up experiment, this paper uses OmegaWave Sport Technology system to track and test the heart rate variability level and central nervous system parameters of 25 elite Chinese women volleyball players who participated in the national adult volleyball training in 2019 and 2020. It is found that the HRV time-domain index of the players under the stimulation of three stages of training load during the winter training in 2020 is determined. Frequency-domain index has significant influence on response stability of central nervous system. In order to further explore the influence of HRV on response stability of central nervous system, a feature classification method based on distance evaluation is proposed for experimental data processing. Through the multimodal human-machine interaction (M-HMI), advanced machine learning is used to promote the cooperative interaction between human and intelligent body. After analysis, SDNN and LF n.u. have a significant impact on the average reaction time. It shows that some indexes tested by the OmegaWave system can reflect the real-time physical function state of athletes sensitively and play an active role in diagnosis of fatigue of athletes' central nervous system. HRV time-domain and frequency-domain indexes, as parameters to evaluate the body functional state of excellent female volleyball players in the preparation process of competition, can sensitively reflect the level of autonomic nerve regulation of athletes in three different load stages.

1. Introduction

Volleyball, as one of the three major balls, has a wide social impact and bears the spirit of the times of the country and nation. In the 1980s, Chinese women's volleyball team reached the historical peak of "five consecutive crowns." In the 21st century, Chinese women's volleyball team won the world championship five times and then made great achievements. Adult women's volleyball team is the frontline team of Chinese women's volleyball team and the basis and guarantee for Chinese women's volleyball team to maintain the world-class team. With the increasing competitive level of modern volleyball and the increase of training load and competition pressure, it is very possible to promote sports fatigue of athletes, which will make them unable to fulfill the technical and tactical requirements laid out by coaches in training and competition and ultimately lead to the decline

of competition ability and affect the effect of training and competition [1–3]. Therefore, how to quickly and accurately monitor the athlete's body function and the functional level of central nervous system and timely understand the athlete's functional state has become an urgent problem for high-level sports teams [4, 5].

Due to the change of external load, a series of physiological and biochemical changes will inevitably take place in athletes. These changes can be counted and measured by detecting physiological and biochemical related indexes. Among the many physiological indexes for measuring sports load, heart rate variability (HRV) is a method of measuring biofeedback that is receiving increasing attention, mainly due to its high technical availability and portability [6, 7]. Exercise load is positively and negatively related to heart rate variability [8]. Increased exercise load leads to inconsistency of ANS functions, which negatively affects HRV [8, 9]. When

an athlete responds to a load intensity, the SNS is activated to allow the athlete to respond appropriately to the sporting needs. However, when an athlete experiences a load stimulus that exceeds his or her current level of acceptance, the SNS response increases [10]. Increased activation of SNS decreases the function of the vagus, which is critical to maximizing and lowering the heart rate during RSA [11]. Therefore, a decrease or even disappearance of vagus nerve tension can reduce the rhythm and have a negative impact on heart rate variability. Therefore, heart rate variability (HRV) is considered by many researchers as a marker of homeostasis [12] and is widely used as an indicator of training adaptability in sports environment [13–17]. Factors such as training load, type, stage, type of competition, and level of health [18–20] have proved to influence HRV level.

In recent years, great advances have been made in computational intelligence and machine learning methods, which have driven the deployment of neural networks and intelligent systems in many life scenarios and industrial fields. Multimodal human-computer interaction system (M-HMI) mainly includes EEG signal and ECG signal. The traditional single-mode human-computer interaction system has been unable to meet the actual needs due to its few task categories, so the multimodal human-computer interaction system came into being. At the same time, other types of interaction are also gradually applied to the human-computer interaction system. Based on this, a multimodal human-machine interaction system based on distance assessment feature classification was constructed by combining ECG signal and EEG signal, which solved the problems of hardware and software platform construction and signal synchronization, and an effective feature classification processing method was proposed.

This study utilizes OmegaWave Sport Technology system. Using omegawave sport technology system, this study analyzes 13 (2019) and 12 (2020) teams participating in China Volleyball Super League. This paper studies the changes of physical function and central nervous system function level of athletes in different intensity training. According to the changes of physiological indexes of athletes after different load training, the experimental results provide a reference for load arrangement of adult women volleyball training course. At the same time, it can improve the training efficiency of women volleyball team. It provides theoretical basis for precompetition training of women volleyball players.

2. Proposed Feature Classification Method Based on Distance Evaluation

The goal of cluster analysis is to collect data and classify them on a similar basis. In this paper, XGBoost is used to model and classify the experimental results, and the possible overfitting caused by XGBoost is improved by calculating the coefficient weight of the characteristic through improved distance evaluation method.

XGBoost can adapt itself to learn a certain number of samples with certain characteristics. However, XGBoost's strong learning ability often results in overfitting, which affects the classification results of samples. Therefore, additional conditions are required to limit the learning ability of the XGBoost model. The improved distance evaluation algorithm can calculate the differences between classes of samples to determine the influence of various characteristics in the sample capacity and the size characteristics of the differences between the categories. The improved distance evaluation method is used to obtain the impact weight of sample features in the model tree, so as to improve the overfitting learning of XGBoost and the classification effect of XGBoost.

XGBoost can quickly classify sample characteristics by decision tree classification. Compared with other decision tree models, this model has faster accuracy and calculation speed. Combining the weight between features obtained by improved distance evaluation algorithm with the weight parameters in XGBoost classification tree model can ensure the correlation between data and prevent the overfitting in the XGBoost method. The parameters in the model are calculated as follows.

Enter characteristic dataset, in which the sample category is Y , the number of samples in each category is N , and the number of features in each sample is M :

$$D = \{X_{y,n,m}, Y\}, \quad (1)$$

where $X_{y,n,m}$ denotes the n th sample in the class y of a dataset that contains m features and y is the category of the corresponding sample.

Standard deviation between data within the calculation coefficient:

$$\delta_{y,m} = \sqrt{\frac{\sum_{n=1}^N (X_{y,n,m} - u_{y,m})^2}{N - 1}}, \quad (2)$$

where $u_{y,m}$ represents the average value of the data within the m th feature in the dataset:

$$u_{y,m} = \frac{1}{N} \sum_{n=1}^N X_{y,n,m}. \quad (3)$$

The standard deviation within the coefficient can be obtained:

$$clt_m^{\text{inner}} = \frac{1}{Y} \sum_{y=1}^Y \delta_{y,m}. \quad (4)$$

There are differences between different categories of data, where y, c indicate that data belong to different categories:

$$f_m^{\text{inner}} = \frac{\max(clt_{y,m}^{\text{inner}})}{\min(clt_{c,m}^{\text{inner}})}. \quad (5)$$

Then, the data differences between features are calculated. First, the standard deviation between features is calculated:

$$\tau_{y,m} = \sqrt{\frac{\sum_{i=1}^{M-1} (cd_{n,r,y,m} - d_{y,m})^2}{N(N-1) - 1}},$$

$$cd_{n,r,y,m} = |X_{y,n,m} - X_{y,r,m}|, \quad (6)$$

$$d_{y,m} = \frac{1}{N(N-1)} \sum_{n,r=1}^N cd_{n,r,y,m}.$$

Calculate the average of standard deviations between features:

$$clt_m^{\text{outer}} = \frac{\sum_{y,c=1}^Y (\tau_{y,m} - \tau_{c,m})^2}{Y(Y-1)}. \quad (7)$$

Differences between features:

$$f_m^{\text{outer}} = \frac{\max(clt_{y,m}^{\text{outer}})}{\min(clt_{c,m}^{\text{outer}})}. \quad (8)$$

The distance weight coefficient between the coefficients can be calculated as

$$\eta_m = \frac{1}{(f_m^{\text{inner}} / \max(f_m^{\text{inner}})) + (f_m^{\text{outer}} / \max(f_m^{\text{outer}}))} \cdot \frac{clt_m^{\text{outer}}}{clt_m^{\text{inner}}}. \quad (9)$$

3. Numerical Experiments

3.1. Experimental Conditions. During the test, the OmegaWave Sport Technology system developed by a LLC Corporation was used to sample the athletes at three stages of different training loads during the winter training. The OmegaWave Sport Technology system developed by LLC Corporation of America is used in the test process of this paper. Athletes are sampled at three stages of different training loads during winter training. The TGAM module includes a TGAT chip, which is a highly integrated EEG sensor. It reads human brain signals using dry electrodes, filters out disturbances from ambient noise, and converts the detected brain signals into digital signals. The device automatically detects abnormal contact conditions and filters out electrical noise and 50/60 Hz AC interference. Bluetooth sensor is used to collect EEG information of athletes in different states, and then it is transmitted back through Bluetooth, electrode pieces are used to collect athletes' ECG information, and the signal collector amplifies the collected signal and stores it on the computer after synchronization. The OmegaWave Sport Technology system is shown in Figure 1.

The TGAM module can directly connect the dry contact points, unlike the wet sensor used in traditional medicine, which requires conductive adhesive, and the single EEG channel has three contact points: EEG (EEG acquisition point), REF (reference point), and GND (ground point). The collected signal is shown in Figure 2.

The physical quantities were measured as follows:

- (1) Measuring original brain wave signal.
- (2) Processing and outputting α , β isoencephalic band data.
- (3) Processing and outputting Neurosky's eSense degree of concentration and relaxation index and other data to be developed in the future.

The research object of this study is to study the physical health of 13 (2019) and 12 (2020) teams of Chinese Women's Volleyball Super League in 2019 and 2020 during the training period of National Adult Women's Volleyball Team, each team having 5 athletes, a total of 125 people. The details are shown in Table 1.

This research adopts preexperiment and postexperiment design modes to test the real-time performance of all the tested athletes who participated in the training for two years after the specific training class and to track and monitor the influence of training load arrangement on the HRV level of the athletes at different training stages during the winter training period. Before winter training, the characteristics of training modes and training loads of athletes in three stages, i.e. early stage, middle stage, and late stage of winter training, are classified. The coaches select a training session in this training stage for testing after a training session and the testing time is within one hour after the end of training. Each participant was tested three times during the whole winter training period, during which HRV index data were collected strictly according to the test process.

As can be seen from Figure 3, in the early stage of winter training in 2019, the basic technical and tactical training is mainly multiball training. Physical training focuses on waist and abdomen strength, lower limb strength, explosive strength, and upper limb strength training. Simulation games mainly focus on group tactical explanation and training. The rest adjustment is based on stretching and active relaxation. In the middle of winter training, the training mode has a certain change, and the basic technical and tactical training is mainly series training. Physical training is mainly based on speed training methods such as step running, slope running, and lateral movement. The simulation competition is mainly based on the actual combat competition of the whole team. The rest adjustment is based on stretching and active relaxation. The latter part of winter training is the combination of the previous two training modes, the basic technical and tactical training to the overall series and confrontation training. Physical training is based on sensitive and flexible training methods. The simulation competition is mainly based on the actual combat competition of the whole team. The training mode for 2020 is shown in Figure 4. The two training modes are the same in structure, but slightly different in time.

3.2. Real-Time Performance Testing Process of Athletes. In the OmegaWave real-time functional test, the basic potential at rest has been identified as an indicator of the functional state and adaptive reserve level of the central nervous system. For healthy people, the significance of the fourth-order resting potential is as follows: less than -30 mV—very low level, -29 to -1 mV—low level, 0 to 46 mV—best level, and greater

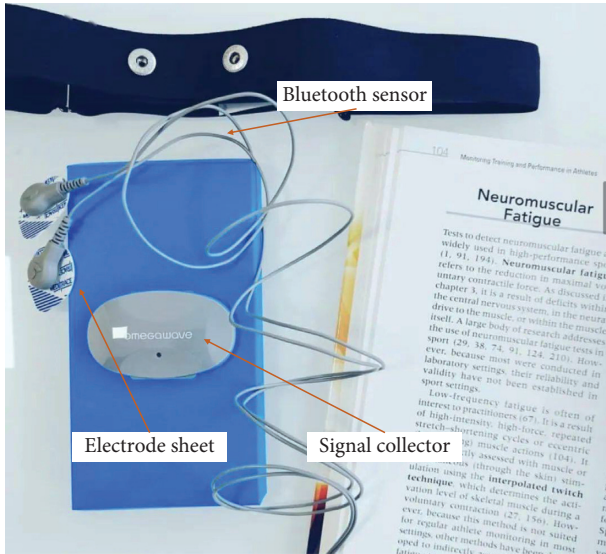


FIGURE 1: OmegaWave Sport Technology system.

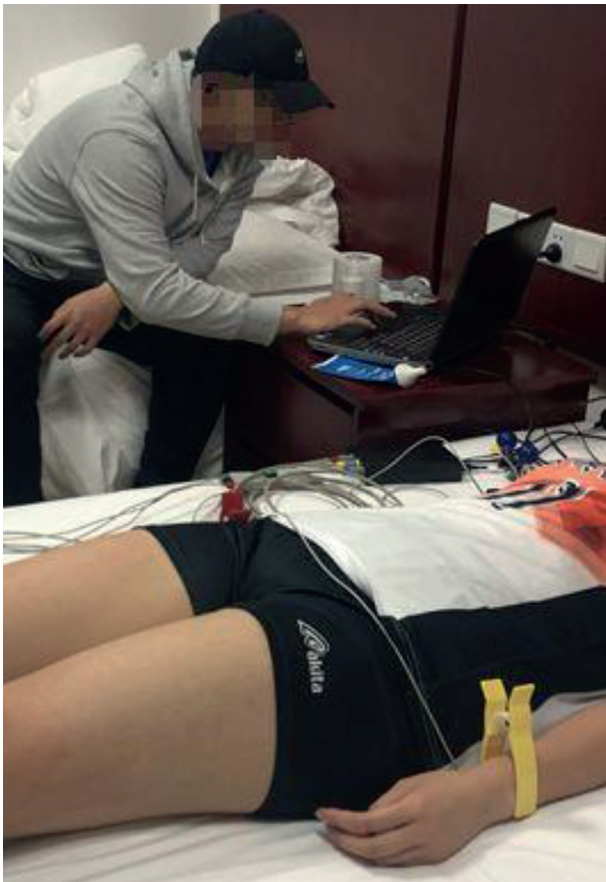


FIGURE 2: Field data acquisition picture.

than 47 mV—high level. The basic test procedure is as follows:

- (1) Enter the basic information such as age, height, weight, and sports grade into OmegaWave system in advance.

TABLE 1: General situation of research objects.

Gender	Age (years)	Training years (years)	Master (person)	Level 1 (people)
Female	21.3 ± 3.1	7.9 ± 2.7	101	24

- (2) Check the basic information of athletes and explain the testing process and requirements before starting the test formally.
- (3) Players lie down in a quiet and comfortable environment.
- (4) The tester wipes the electrode with an alcohol cotton ball at the position where the electrode is affixed according to the operation requirements and then clips the electrode.
- (5) Confirm that brain lead, chest lead, and limb lead are connected to computer normally and the athletes have no uncomfortable reaction.
- (6) Athletes remain relaxed and then begin testing. There are differences in the time of reaching steady state among different athletes and the time of testing. The whole data collection process is completed in about 15 minutes.
- (7) The test data are collected and the reaction time test is carried out after the relevant lead is removed. The test indexes are shown in Table 2. They mainly include the standard deviation of NN interval (SDNN), the mean square deviation of adjacent NN interval difference (RMSSD), and the standard deviation of adjacent NN interval difference (SDSD).

The testing process is shown in Figure 5. ECG and EEG signals of the tested personnel are collected first, amplified and stored, and then preprocessed for feature classification to obtain correlation coefficients between features.

4. Test Results

4.1. Variation Characteristics of HRV Time-Domain Indexes of Elite Female Volleyball Players in Different Training Load Stages. Heart rate variability (HRV) index can reflect the activity of the autonomic nervous system, the tension of sympathetic nerve and vagus nerve, and the influence of autonomic nervous system on athletes' heart rate and can reveal the more complex change law of heart rate. There is a certain correlation between exercise load and heart rate variability index. In previous studies, SDNN decreased significantly after exercise, and it is significantly negatively correlated with the increase of biochemical index BLA reflecting load intensity.

Figures 6 and 7 show the influence of different training load stimulation on HRV time-domain indexes of elite female volleyball players; as can be seen from the figures, SDNN is the most sensitive index, so it can be used to reflect the actual situation of athletes. Table 3 shows that SDNN of athletes increased gradually with the exercise time. During the winter training in 2020, athletes gradually entered a state of tension, and the ability of autonomic nervous system to

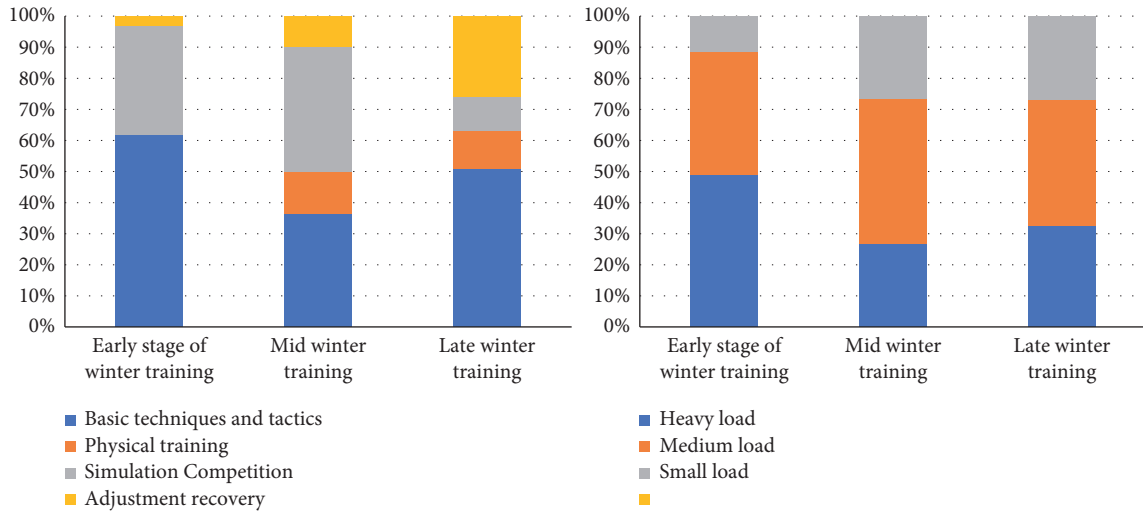


FIGURE 3: Proportion of winter training course plan and training load in 2019.

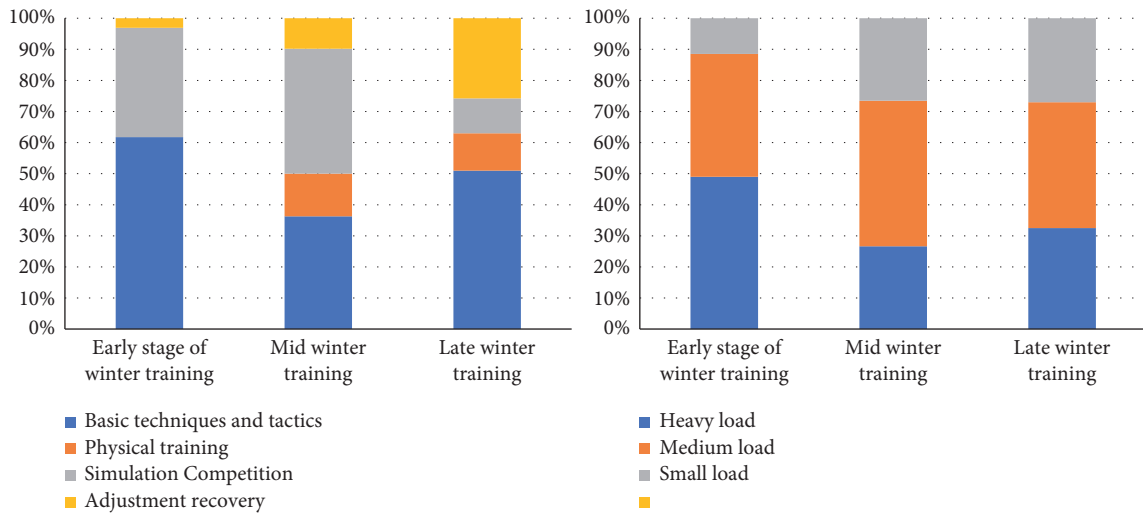


FIGURE 4: Proportion of winter training course plan and training load in 2020.

TABLE 2: List of test indicators.

	Test indicators					
Time-domain indicators	SDNN	RMSSD	SDSD		PNN50	
Frequency-domain index	LF/HF	HF	HF n.u.		LF	LF n.u. VLF
Indicators of central nervous system	Quiet potential value	Average reaction time	Functional level index of sensory motor nervous system		Reaction stability index	Sensory motor development potential index Tension index

regulate heart rate decreased gradually; the change of load intensity and the increase of psychological pressure near the game may be the reasons for this phenomenon. At the same time, from the arrangement of training load in 2020, it can be found that in the test process of the initial stage of winter training, the daily training of volleyball players is mainly

based on basic technical training and physical fitness training, the arrangement of simulated competition is less, and the load intensity borne by athletes is also low. In the middle of winter training, targeted training such as attack and defense confrontation has increased; especially, when coaches carry out multigroup attack and defense series

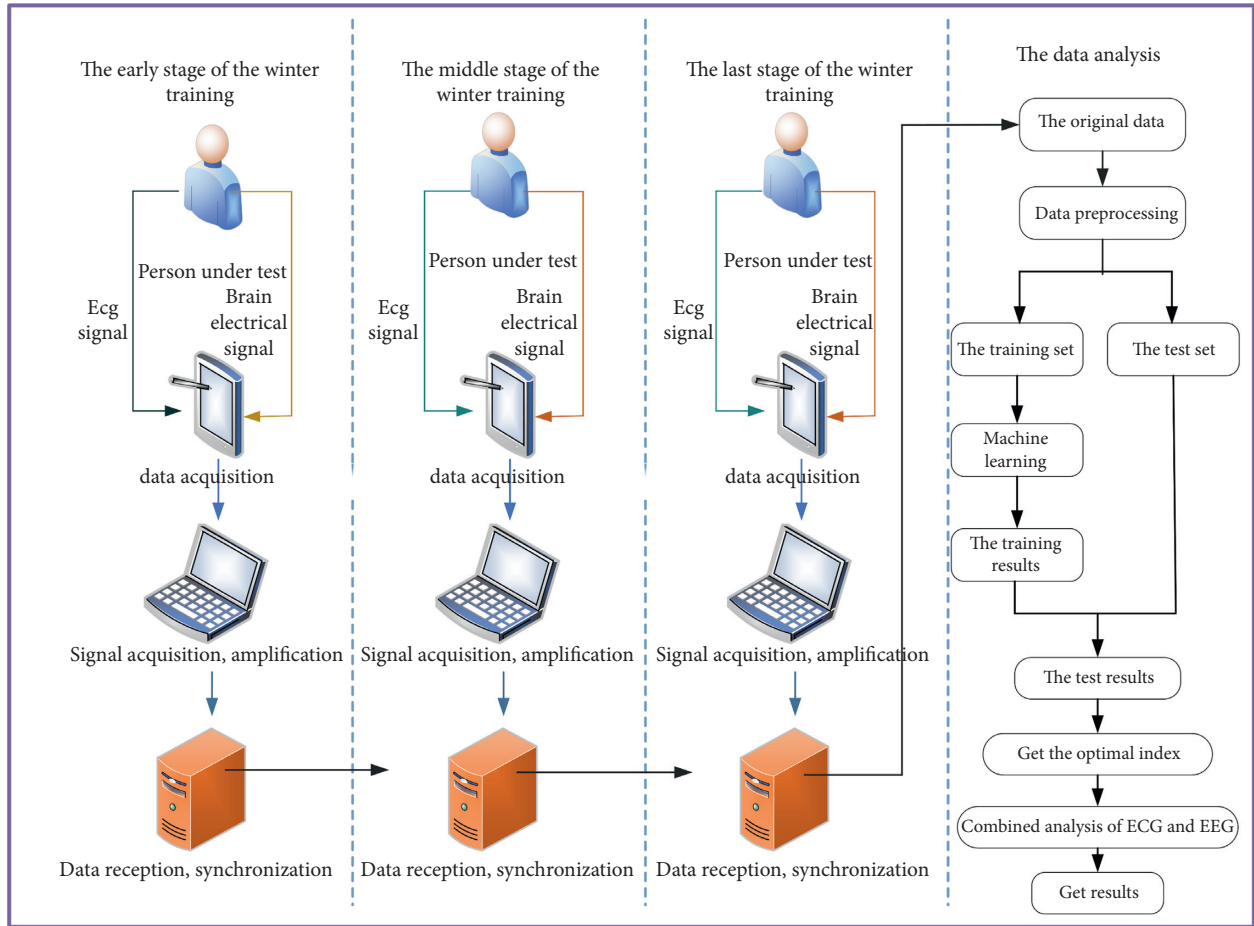


FIGURE 5: The flowchart of the method.

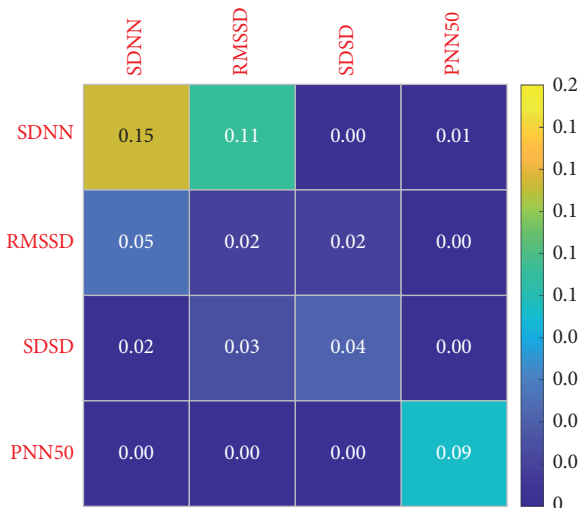


FIGURE 6: Change sensitivity of HRV time-domain indexes of women volleyball players in different load stages in 2019.

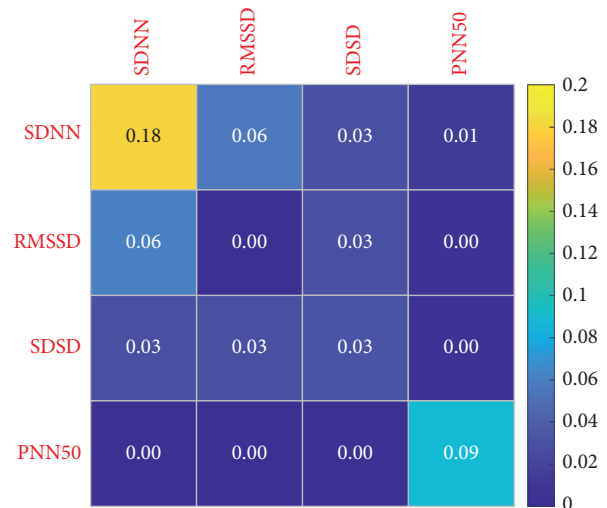


FIGURE 7: Change sensitivity of HRV time-domain indexes of women volleyball players in different load stages in 2020.

training for athletes, the number of touch times, moving distance, and take-off times have increased sharply in a short time, so that the load intensity borne by athletes is greater than that in the early training stage of winter training. In the late winter training period, the athletes' load intensity is also

higher and higher. The proportion of simulation competitions is also increasing. This also causes athletes to increase the probability of sports fatigue. Therefore, the proportion of adjustment and recovery is increasing in the arrangement of training plans.

TABLE 3: Changes of HRV time-domain indexes of women volleyball players at different load stages.

	SDNN	RMSSD	SDSD	PNN50
The early stage of the winter training	52.31 ± 2.24	62.12 ± 3.45	81.17 ± 4.12	18.96 ± 2.94
Middle of winter training	55.93 ± 2.64	66.22 ± 3.56	85.91 ± 4.95	21.86 ± 1.16
Late winter training	69.67 ± 3.69	81.34 ± 4.43	103.36 ± 5.91	22.54 ± 1.22
<i>F</i>	2.268	1.414	1.126	0.574
<i>P</i>	0.114	0.252	0.332	0.568

4.2. *Variation Characteristics of HRV Frequency-Domain Indexes of Elite Female Volleyball Players in Different Training Load Stages.* Figures 8 and 9 show the variation characteristics of HRV frequency-domain indicators. As can be seen from the figures, in different training load stages of elite women volleyball players, LF n.u. indicator is the most sensitive and can therefore be used to reflect the actual situation of athletes. As can be seen from Table 4, it first decreases and then increases as the training phase progresses.

This indicates that in the middle stage of winter training, the athletes' vagal tension decreases, their cardiac oxygen consumption increases, and their heart rate recovery slows down after exercise. It may be that the athletes' physical function state at this stage is not fully recovered, or they have not adapted to the exercise load stimulation at this stage. However, in the late stage of winter training, the athletes' training load changes and the proportion of recovery adjustment increases, and their physical function level also recovers. It has exceeded the level of early winter training, or it has not adapted to the exercise load stimulation at the stage, and in the later stage of winter training, with the change of athletes' training load and the increase of recovery adjustment proportion, the level of physical function also recovers, but it does not reach the level at the early stage of winter training. Through the analysis of the proportion of the training plan in the winter training stage in 2020, it can be found that the training focus in the early stage of winter training is mainly on the strengthening basic techniques and tactics. In the middle stage of winter training, the proportion of simulated competitions of each team gradually increases, and the proportion of basic techniques and tactics training and physical training courses gradually decreases. With the winter training reaching the final stage, athletes will have a certain accumulation of sports fatigue and hidden dangers of injuries. In the later stage of winter training, the training plan mainly focuses on simulated competition and recovery adjustment. On the one hand, it tests the effect of technical and tactical training and prepares for future competitions.

4.3. *Variation Characteristics of Central Nervous System Indexes of Elite Female Volleyball Players in Different Training Load Stages.* Figures 10 and 11 show the variation characteristics of HRV frequency-domain indexes of elite female volleyball players at different training load stages. In the test results of this study, the tension index in the later stage of winter training in 2019 is significantly higher than that in the early stage of winter training. As can be seen from Table 5, the reaction time of athletes decreases gradually with training. This phenomenon shows that with the adjustment of winter training load intensity and load, the athletes' stress

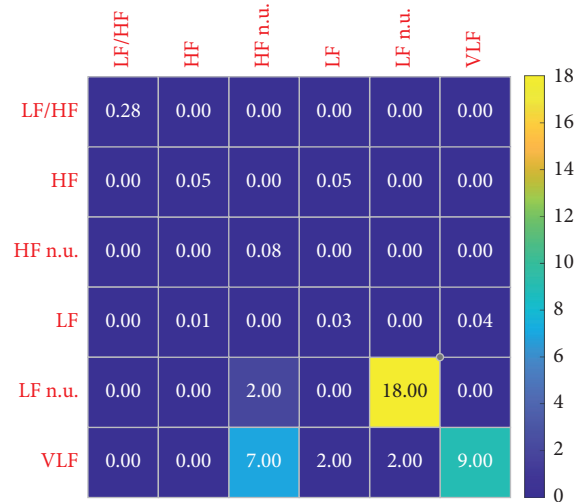


FIGURE 8: Change sensitivity of HRV frequency-domain index of women volleyball players in different load stages in 2019.



FIGURE 9: Change sensitivity of HRV frequency-domain index of women volleyball players in different load stages in 2020.

response to load stimulation reaches the peak in the later stage. In the later stage of winter training, the proportion of each sports team has not been greatly adjusted. In 2020, the reaction stability index in the middle of winter training was significantly lower than that in the early stage of winter training, and it recovered in the later stage of winter training. This shows that in the middle stage of winter training, with the increase of simulated competition, the stimulation of

TABLE 4: Changes of HRV frequency-domain indexes of female volleyball players at different load stages.

	LF/HF	HF	HF n.u.	LF	LF n.u.	VLF
The early stage of the winter training	0.85 ± 1.45	836.46 ± 32.57	69.67 ± 2.72	293.64 ± 30.67	30.33 ± 2.72	92.59 ± 3.78
Middle of winter training	0.42 ± 0.30	746.62 ± 35.09	73.38 ± 3.40	293.58 ± 30.55	26.62 ± 3.40	90.79 ± 3.16
Late winter training	0.99 ± 2.18	1300.62 ± 37.56	67.49 ± 1.62	989.18 ± 12.46	32.51 ± 9.62	101.69 ± 6.74
<i>F</i>	0.770	1.628	0.492	2.394	0.492	0.134
<i>P</i>	0.468	0.205	0.614	0.100	0.614	0.875

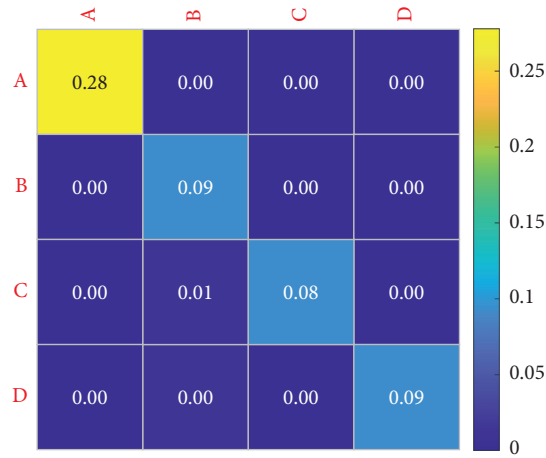


FIGURE 10: Sensitivity of central nervous system indexes of women volleyball players at different load stages in 2019.

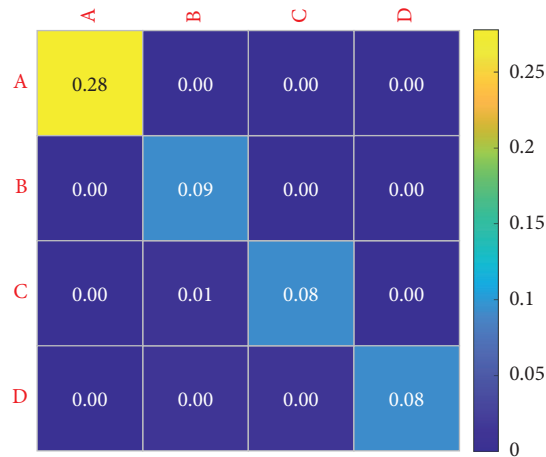


FIGURE 11: Sensitivity of changes in central nervous system indexes of women volleyball players at different load stages in 2020. A: mean response time; B: sensory motor nervous system function level index; C: response stability index; D: sensorimotor function development potential index.

TABLE 5: Changes of central nervous system indexes of female volleyball players at different loading stages.

	A	B	C	D
The early stage of the winter training	-1.03 ± 0.016	4.79 ± 0.21	3.78 ± 0.42	117.31 ± 3.5
Middle of winter training	0.64 ± 0.042	4.64 ± 0.21	3.42 ± 0.49	93.50 ± 1.94
Late winter training	-0.52 ± 0.09	4.72 ± 0.27	3.66 ± 0.52	118.43 ± 7.07
<i>F</i>	0.075	2.627	3.757	0.517
<i>P</i>	0.927	0.079	0.026	0.597

athletes' load intensity on the pivot nerve is significantly higher than that in the early stage of winter training. Because the simulated competition is the actual competition of the

whole team, this training method requires athletes' attention to be highly concentrated compared with ordinary tactical training. In the middle stage of winter training, the recovery

mode of each team is mainly physical relaxation. The lack of recovery of the central nervous system may also be one of the reasons for the significant decline of the response stability index of the central nervous system at this stage.

5. Conclusion

- (1) The influence of different training loads on the results of real-time functional state test of female volleyball players is significant. Each index has different characteristics between the two experiments. Therefore, real-time functional state test plays a positive role in the diagnosis of fatigue of athletes' central nervous system.
- (2) HRV time-domain, frequency-domain, and central nervous system parameters are used to evaluate the physical function of elite women volleyball players in the process of intensive training and preparation before the competition. From the results, it can be seen that SDNN in time-domain index is related to the average reaction time of central nervous system.
- (3) HRV time-domain index SDNN and frequency-domain index LF n.u. have a significant impact on the average reaction time. The specific impact law needs to be further explored.

Data Availability

The data used to support the findings of this study have not been made available because they are confidential.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Liao Ludi wrote the paper. Li Jianying performed the experiments. All authors have read and approved the final manuscript.

References

- [1] Z. Hu, C. Lv, P. Hang, C. Huang, and Y. Xing, "Data-driven estimation of driver attention using calibration-free eye gaze and scene features," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 2, pp. 1800–1808, 2022.
- [2] M. K. Beauchamp, R. H. Harvey, and P. H. Beauchamp, "An integrated biofeedback and psychological skills training program for Canada's Olympic short-track speedskating team," *Journal of Clinical Sport Psychology*, vol. 6, no. 1, pp. 67–84, 2012.
- [3] M. Paul and K. Garg, "The effect of heart rate variability biofeedback on performance psychology of basketball players," *Applied Psychophysiology and Biofeedback*, vol. 37, no. 2, pp. 131–144, 2012.
- [4] Z. Hu, Y. Xing, C. Lv, P. Hang, and J. Liu, "Deep convolutional neural network-based Bernoulli heatmap for head pose estimation," *Neurocomputing*, vol. 436, pp. 198–209, 2021.
- [5] E. P. Paula Jr., D. S. Paza, G. C. Pierozan, and J. F. Stefanello, "Heart rate variability and emotional states in basketball players," *Journal of Exercise Physiology Online*, vol. 19, no. 6, pp. 111–122, 2016.
- [6] I. Z. Khazan, *Clinical Handbook of Biofeedback: A Step-by-step Guide for Training and Practice with Mindfulness*, Wiley-Blackwell, Malden, MA, 2016.
- [7] J. Fatissou, V. Oswald, and F. Lalonde, "Influence diagram of physiological and environmental factors affecting heart rate variability: an extended literature overview," *Heart International*, vol. 11, p. 32, 2016.
- [8] A. A. Flatt and D. Howells, "Effects of varying training load on heart rate variability and running performance among an olympic rugby sevens team," *Journal of Science and Medicine in Sport*, vol. 22, no. 2, 2019.
- [9] A. A. Flatt, M. R. Esco, J. R. Allen et al., "Cardiac-autonomic responses to in-season training among division-1 college football players," *The Journal of Strength & Conditioning Research*, vol. 34, no. 6, pp. 1649–1656, 2020.
- [10] D. Atlaoui, V. Pichot, L. Lacoste, F. Barale, J.-R. Lacour, and J.-C. Chatard, "Heart rate variability, training variation and performance in elite swimmers," *International Journal of Sports Medicine*, vol. 28, no. 5, pp. 394–400, 2007.
- [11] L. Schmitt, J. Regnard, A. Parmentier et al., "Typology of "fatigue" by heart rate variability analysis in elite nordic-skiers," *International Journal of Sports Medicine*, vol. 36, no. 12, pp. 999–1007, 2015.
- [12] Y. Le Meur, A. Pichon, K. Schaal et al., "Evidence of parasympathetic hyperactivity in functionally overreached athletes," *Medicine & Science in Sports & Exercise*, vol. 45, no. 11, pp. 2061–2071, 2013.
- [13] A. A. Flatt, B. Hornikel, and M. R. Esco, "Heart rate variability and psychometric responses to overload and tapering in collegiate sprint-swimmers," *Journal of Science and Medicine in Sport*, vol. 20, no. 6, pp. 606–610, 2017.
- [14] J. Stanley, J. M. Peake, and M. Buchheit, "Cardiac parasympathetic reactivation following exercise: implications for training prescription," *Sports Medicine*, vol. 43, no. 12, pp. 1259–1277, 2013.
- [15] D. J. Plews, P. B. Laursen, J. Stanley, A. E. Kilding, and M. Buchheit, "Training adaptation and heart rate variability in elite endurance athletes: opening the door to effective monitoring," *Sports Medicine*, vol. 43, no. 9, pp. 773–781, 2013.
- [16] W. Zhao, Z. Wang, W. Cai, and Q. Zhang, "Multiscale Inverted Residual Convolutional Neural Network for Intelligent Diagnosis of Bearings under Variable Load Condition," *Measurement*, vol. 188, Article ID 110511, 2022.
- [17] A. L. Wheat and K. T. Larkin, "Biofeedback of heart rate variability and related physiology: a critical review," *Applied Psychophysiology and Biofeedback*, vol. 35, no. 3, pp. 229–242, 2010.
- [18] Z. Wang, J. Zhou, Y. Lei, and W. Du, "Bearing fault diagnosis method based on adaptive maximum cyclostationarity blind deconvolution," *Mechanical Systems and Signal Processing*, 2021.
- [19] J. I. Lacey and B. C. Lacey, "Some autonomic-central nervous system interrelationships," in *Physiological Correlates of Emotion*, P. Black, Ed., Academic Press, New York, NY, USA, pp. 205–227, 1970.
- [20] B. C. Lacey and J. L. Lacey, "Studies of heart rate and other bodily processes in sensorimotor behavior," in *Cardiovascular Psycho-Physiology: Current Issues in Response Mechanisms, Biofeedback and Methodology*, P. A. Obrist, A. H. Black, J. Brener, and L. V. DiCara, Eds., pp. 538–564, Aldine Transaction, New Brunswick, NJ, 1974.

Research Article

A Safe and Compliant Noncontact Interactive Approach for Wheeled Walking Aid Robot

Donghui Zhao ¹, **Wei Wang**,¹ **Moses Chukwuka Okonkwo**,¹ **Zihao Yang**,^{1,2}
Junyou Yang ¹ and **Houde Liu**³

¹School of Electrical Engineering, Shenyang University of Technology, Shenyang, China

²Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China

³Center for Artificial Intelligence and Robotics, Tsinghua University, Shenzhen, China

Correspondence should be addressed to Donghui Zhao; putongdeyu@126.com

Received 9 January 2022; Revised 16 February 2022; Accepted 25 February 2022; Published 16 March 2022

Academic Editor: Jie Liu

Copyright © 2022 Donghui Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at promptly and accurately detecting falls and drag-to gaits induced by asynchronous human-robot movement speed during assisted walking, a noncontact interactive approach with generality, compliance and safety is proposed in this paper, and is applied to a wheeled walking aid robot. Firstly, the structure and the functions of the wheeled walking aid robot, including gait rehabilitation robot (GRR) and walking aid robot (WAR) are illustrated, and the characteristic futures of falls and the drag-to gait are shown by experiments. To obtain gait information, a multichannel proximity sensor array is developed, and a two-dimensional gait information detection system is established by combining four proximity sensors groups which are installed in the robot chassis. Additionally, a node-iterative fuzzy Petri net algorithm for abnormal gait recognition is proposed by generating the network trigger mechanism using the fuzzy membership function. It integrates the walking intention direction vector by taking gait deviation, frequency, and torso angle as input parameters of the system. Finally, to improve the compliance of the robot during human-robot interaction, a PID_SC controller is designed by integrating the gait speed compensation, which enables the WAR to track human gait closely. Abnormal gait recognition and assisted walking experiments are carried out respectively. Experimental results show that the proposed algorithm can accurately identify abnormal gaits of different groups of users with different walking habits, and the recognition rate of abnormal gait reaches 91.2%. Results also show that the developed method can guarantee safety in human robot interaction because of user gait follow-up accuracy and compliant movements. The noncontact interactive approach can be applied to robots with similar structure for usage in walking assistance and gait rehabilitation.

1. Introduction

Due to an increase in the percentage of the aging population and the growing number of disabled people with lower limb impairment, there is a significant rise in the demand for walking aids or professional care attendants in the daily lives of the affected people. But this increasing demand cannot be sufficiently satisfied due to current shortage in supply [1, 2]. Hence, the development of a robot which can help in walking for both the elderly and disabled during rehabilitation has become a prominent issue in the field of robotics [3].

The compliant control of robots for assisted walking is an important research subject regarding the medical care

needed for weak motion capability user groups. In recent years, researcher globally, have studied this and proposed several methods [4]. JR et al. proposed a robot walking control method based on COR (center of rotation), which enables physically impaired users to compliantly walk by changing the kinematic structure of the system [5]. Jamwal et al. proposed the impedance control method. This control method has small computation and strong robustness, and has been widely used in the compliant control field of WAR. However, when there is external disturbance in the human-robot interaction environment, this control system cannot maintain optimal impedance control throughout the whole process [6]. On this basis, Jiang et al. proposed a shared

control method that takes the robot system and human-robot interaction as inputs, which can assign a degree of control to the user according to different use scenarios [7] (this control method change the main control source depending on the situation to either the user or the robot). Xu et al. proposed a shared control method based on reinforcement learning algorithm, which enables the robot to adapt to the user's personal characteristics and gait, thus making walking assistance more stable [8]. To improve the safety of the user during rehabilitation, Xu et al. proposed a compliant control method based on multisensor fusion technology. This method improved the compliance of the rehabilitation robot, nevertheless, the method has a poor walking intention recognition performance in an environment containing non-Gaussian noise [9]. Han et al. proposed trajectory tracking control method based on PID neural network to enhance the compliance in human-robot interaction [10]. Yan et al. developed a force sensor array to measure human-robot interaction force of the user's upper limbs to calculate required motion intention information. Laser sensors are then used to measure leg distance to predict the walking intention of the user's lower limbs. For compliant robot movements, the above obtained data are fused with Kalman filter algorithm to acquire the user's walking intention speed [11]. Hirata et al. proposed updating the state estimation parameters of the user to realize robot compliant motion control by observing the human-robot physical interaction which supports the user's assisted walking [12]. Song et al. developed an external force observer based on the measurement of motor current and speed. The robot adapts by moving compliantly according to the force applied by the user, but the robot cannot quickly process any instantaneous data like pulls and push caused by the user's fall or other emergency situations, thus increasing the risk of secondary injury during a fall [13, 14].

Although a large number of studies have been carried out on the compliant control of walking rehabilitation robots, extreme situations such as robot induced falls and drag-to gaits which have not been considered in the recognition of abnormal human walking intentions, makes it impossible to effectively guarantee absolute safety in human-robot interaction. Regardless of using robots as a walking aid or in more complex rehabilitation scenarios, ensuring user safety is an important factor to consider during research [15, 16]. If the robot has the ability to detect a fall before the patient reaches the ground thereby enabling its prevention, user safety will be ensured.

Aiming at this kind of safety issue, researchers have proposed some fall detection methods, such as wearable sensor detection [17, 18], visual detection [19, 20], and environmental monitoring [21, 22]. A wrist watch with built-in accelerometer can be used for fall detection by monitoring the amplitude of acceleration with matching support vector machine algorithm [17]. The fall detection algorithm based on the fusion of plantar pressure signal and surface EMG signal achieves an average recognition rate of 91.7% in normal day usage [23]. Huang et al. [24], Ma et al. [25], and Qiu et al. [18] put forth the body posture estimation algorithm based on wearable sensor which can effectively estimate the user's posture, detect abnormal behavior in real

time and monitor the user online. Unfortunately, for the wearable fall recognition methods, special sensors need to be worn in advance to collect and store walking and fall data incidences which increases usage complexity and poor user experience. Consequently, Lee et al. proposed a method to estimate the user's walking state based on visual inspection [19]. Kalman recursive prediction based on real-time measurements of the knee angle actualizes effective abnormal gait monitoring [26]. This visual inspection method has both low cost and ease of use advantages as there is no need to wear any equipment. But the disadvantage is that it can only be used in the environment where sensors are installed. Li et al. proposed a method based on modified zero moment point for fall predictions and used Kinect sensor to monitor user's movements [21]. Di et al. proposed the estimation of the user's center of gravity in real time based on COP-FD algorithm and ZMP (zero moment point) algorithm [27, 28], so as to monitor the user's state online. Yan et al. proposed a human-robot cooperative stability algorithm to measure the walking state of both the robot and user [29]. Wakita et al. have designed a robot with a smart cane to help the elderly and disabled during walking. The concept of "intention direction" was proposed and various sensors are used to detect the user's intention. But the human-robot interface which uses multi-axis sensors is expensive and fragile, and not affordable for a large number of users [30]. In the process of using a wheeled walking aid robot, a neuromuscular disorder in the user's lower limb may lead to walking ability degeneration, abnormal gait, and body imbalance. Developments such as robot induced drag may be observed due to the failure of the patient to follow the robot in time. This generally led to falls causing dangerous secondary injuries. At present, there is no research on these particular robots induced abnormal gait which usually occurs before users fall. Meanwhile, the noncontact interaction mode is more convenient for the users with weak motion capability, which could avoid cumbersome steps, such as repeated wearing and data correction in advance [31, 32]. In particular, for gait rehabilitation training, the noncontact interaction mode enables users to actively master the gait rhythm and gait phase. It helps to promote the active participation of users and improve the effect of gait training [33].

To recognize abnormal gaits accurately, we proposed a node-iteration fuzzy petri net algorithm (NIFPN), which is applied in the gait rehabilitation robot (GRR) and walking aid robot (WAR). Additionally, we developed a compliant PID_SC controller, which can track the user's gaits accurately. On the whole, a noncontact interactive approach which ensures both safety and compliant motion is proposed. The method proposed in this paper has the following innovations:

- (1) A low-cost multichannel proximity sensor is developed to effectively detect real-time gait information by multichannel data fusion. Its unique noncontact design has good generalization characteristics, which means the sensor can be applied to wheeled walking aid robot with similar structure.
- (2) A node-iteration fuzzy petri net (NIFPN) algorithm is proposed to recognize abnormal gait. The

recognition rate is improved by updating nodes for individual behavior differences of users. Compared with the traditional experience-based threshold method, this eliminates involved data precorrection and storage.

- (3) A PID_SC controller is proposed to adequately follow the user's gaits. Human gait speed compensation is introduced in the calculation process of traditional PID controller, which significantly improves the stability and compliance of WAR during assisted walking.

The rest of this paper is as follows: Section 2 analyzes the composition structure and functions of the developed WAR and discussed subsequent preliminary experiment processes in details. Section 3 introduces the abnormal gait recognition data extraction method and proposes the NIFPN algorithm and PID_SC controller. Section 4 and 5, respectively, describes the comprehensive experiment and discusses related conclusion.

2. Materials

Our laboratory has developed two wheeled walking aid robots: gait rehabilitation robot (GRR) and walking aid robot (WAR), as shown in Figure 1. WAR is specially made for assisting the elderly and patients with walking disabilities, while GRR mainly serves medial gait rehabilitation purposes. WAR will be used for the experimental demonstration of the above stated method in this paper.

2.1. Walking Aid Robot. WAR is composed of an operation interface, pressure support plate and gait information detection platform. Four pressure sensors are embedded in the pressure support plate to obtain the direction intention [34]. To further improve the accuracy of information acquisition, we established a gait information detection platform using the multichannel proximity sensor, as shown in Figure 2. The multichannel proximity sensor developed is shown in Figure 2(a), which mainly consists of the laser distance measurement sensor VL6180X, embedded microcontroller SH74552, and CAN transceiver SN65HVD230. In placing two sensors in the front side of the gait information detection platform, we measure the distance between the legs as they swing back and forth. Two other sensors are installed on both sides of the platform to measure the lateral distances of both legs, as shown in Figure 2(b). Distance data from the multichannel proximity sensor is transmitted to the data acquisition circuit, as shown in Figure 2(c), where they are integrated and sent to a microcomputer. Received data are converted from CAN communication protocol to serial communication protocol by the MCU before being sent to the PC. The PC controls the movement of WAR based on the movement status of both legs.

2.2. Preliminary Preparation Experiment. To effectively detect the abnormal gait characteristics of people with lower limb disabilities. Multiple directional walking experiments

and long-distance linear walking experiments were conducted respectively, as shown in Figure 3. More intuitive insight of the user's gaits is gotten by installing a pressure sensor array in the user's shoes, which effectively reflects the center of gravity and gait swing phase of the user's foot. To illustrate the directional interaction between user and robot, we take the forward movement of the user as the front cardinal direction with respect to the robot. Front, right-front and left-front walking directions are shown in Figures 3(a), 3(b), and 3(c), respectively, and the corresponding gait information are shown in Figures 3(d), 3(e), and 3(f). In the right-front walking direction as shown in Figures 3(b) and 3(e), the user's body pressure is concentrated on his right foot, which means that the person intends to initiate a forward right walking movement and to appropriately respond to this, WAR makes a right-front movement. At this time, if the user cannot swing the left foot quickly, a collision may occur or the robot will drag the user towards the corresponding direction. Also, in Figure 3(c), the center of gravity of the user is extremely inclined to the left leg, causing the right foot to easily collide with the side of the robot as the user is dragged along. Falls and induced drag-to gait may not only occur in the already discussed direction but also can occur and be detected as well in all directional movement of the user and robot. Through experiments, it can be observed that when the intention direction line of the user on the co-ordinate plane has a large degree of deviation with respect to the running direction of the robot, abnormal gaits will be induced if not adjusted in time. In practical usage, because of impairment in the lower limb muscles and nerves, or fatigue caused by long usage of the robot, it often happens that the user fails to keep up with the speed of the robot. Users with extreme conditions may fall to the ground because of their inability to self-adjust their gait speed.

3. Method

3.1. System Block Diagram. The WAR provides two operation modules, namely, assisted walking module and abnormal gait recognition module. As shown in Figure 4, an active compliant control method is introduced according to the following: first, the multichannel proximity sensor and pressure sensor on the robot collects the user's gait information and the forearm pressure information respectively. In assisted walking module, the gait information is passed through the data processing to complete the recognition of the user's walking intention. Information from the center position of the user's body is inputted to the PID_SC controller which outputs the driving speed used for the control of WAR during assisted walking.

For the abnormal gait recognition module, information obtained from the multichannel proximity sensor and pressure sensor which includes inclination angle, walking intention deviation angle and frequency serve as the input parameters of NIFPN algorithm. After these parameters are inputted to the algorithm, involved calculations and gait evaluations are done. The algorithm involves node updates which are essential for the optimization of abnormal gait

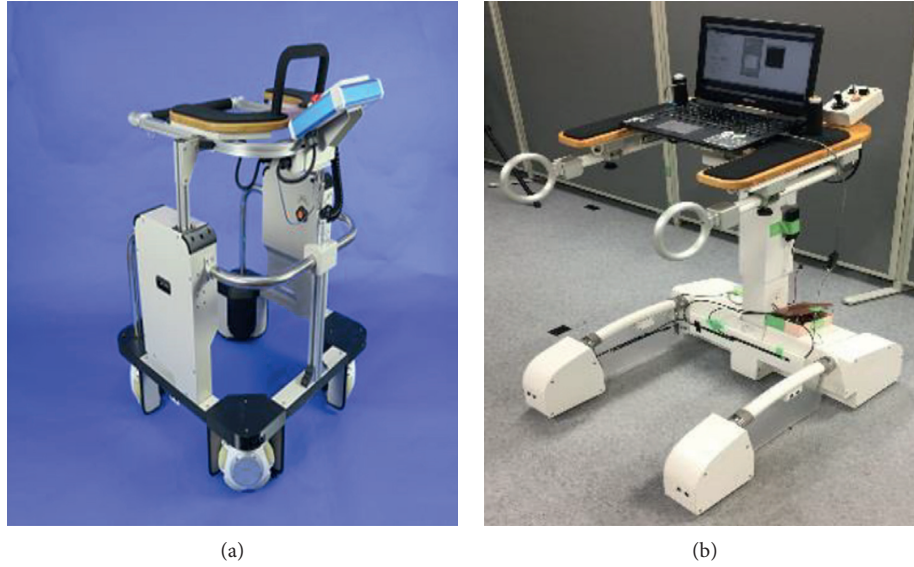


FIGURE 1: Wheeled walking aid robots. (a) GRR. (b) WAR.

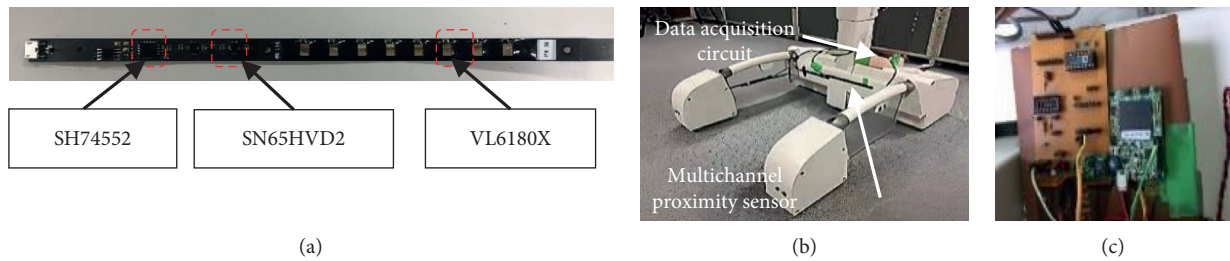


FIGURE 2: Gait information detection platform. (a) Multichannel proximity sensor. (b) Sensor installation. (c) Data acquisition circuit.

recognition for individual users. When an abnormal gait is detected, the robot immediately brakes to prevent dragging the user. Finally, the driving speed output from the PID_SC controller is combined with the emergency braking indicator signal from abnormal gait recognition results to control the movement of WAR

3.2. Walking Intention Recognition and Parameter Extraction. When interacting with robots, walking direction intention and gait information directly reflects movement state.

- (1) Gait information recognition: firstly, error data exceeding maximum distance of multichannel proximity sensor are excluded during the whole calculation. The mean value of the effective data obtain from the multichannel proximity sensor is taken as the observation value. Then, the relative distance between the foot contour and the robot is estimated based on the Kalman filter. The communication frequency of the whole system is 10 Hz. Figure 5 shows the proximity data of the left and right feet during normal and restrained (user suffers an impairment on the left foot) walking. The vertical axis shows the distance measured by the proximity sensor, and the horizontal axis reflects time. For

restrained walking (sensor data shown in Figure 5(a)), the blue signal line shows the position of the right foot with respect to the robot which is considered to support the body weight of the user because it has a shorter displacement. The red signal indicates a larger displacement of the left foot with respect to the robot. In this experiment, if the left foot is slow the user might not be able to follow or keep up with the speed of the robot.

Next, we propose a method to extract abnormal gait parameters and a coordinate system is established with its origin at the ground level of the left front corner of the base of the robot. The height of the robot given as h_s . The coordinate of the force point of the combined upper limb pressure on the pressure support plate is $P_d(x_d, y_d, h_s)$, and the combined vector U directly reflects the inclination degree of the torso, and thus serves as a safety indicator during assisted walking [35]. By selecting the upper left corner pressure sensor as the point of origin and the four sensor values as p_{-fl} , p_{-fr} , p_{-bl} , and p_{-br} , the walking intention direction vector force $F(f_d, \theta)$ is obtain based on the distance type fuzzy inference algorithm [34]. The Euclidean distance between point $P_d(x_d, y_d)$ and the center point $P_s(x_s, y_s)$ of

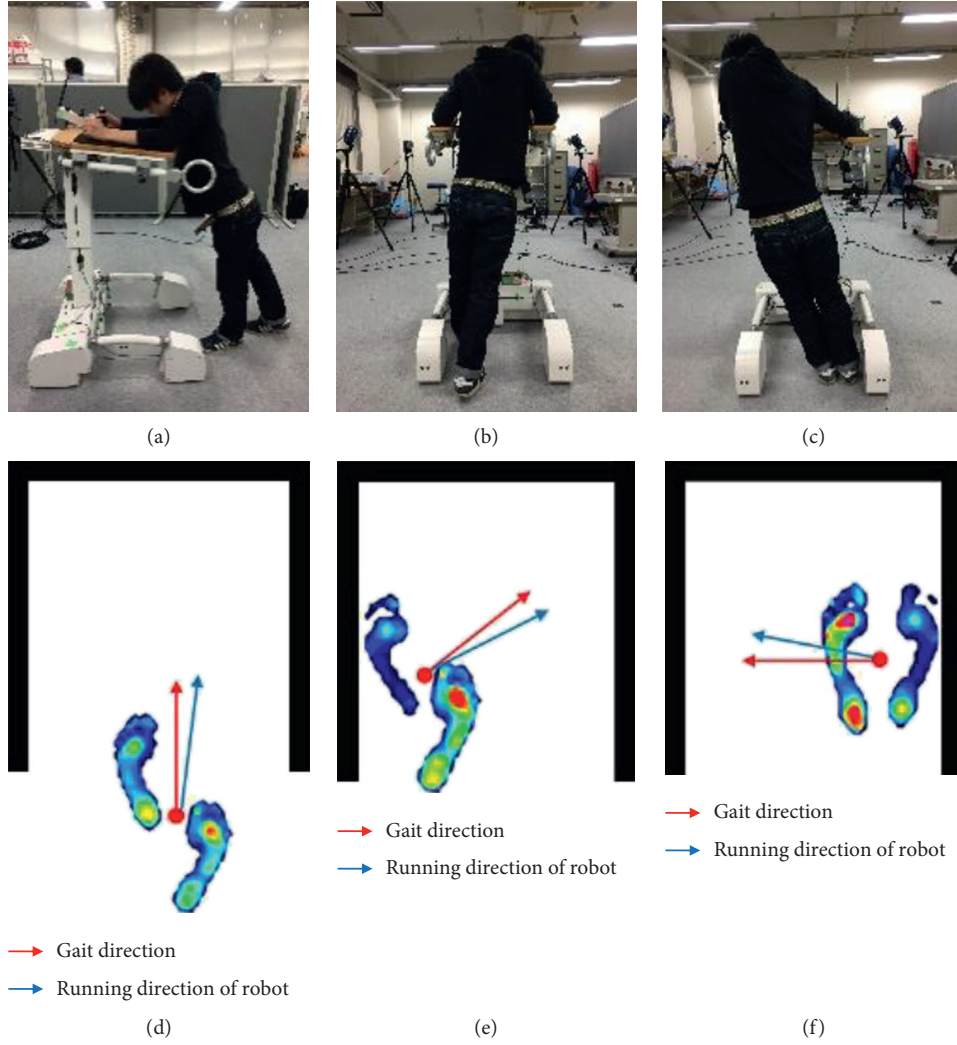


FIGURE 3: Experiment of multimodal assisted walk. (a) Front. (b) Right-front. (c) Left-front. (d) Forward gait. (e) Right front gait. (f) Left front gait.

both feet reflects the deviation degree between the running direction of the robot and the actual foot-step position. Again, in the case of abnormal gait, the user's lower limb cannot properly indicate the user's directional intentions which results in large driving angle and linear distance deviation. Three input parameters of the abnormal gait recognition system are as follows:

- (2) Torso inclination angle θ_s :

$$\theta_s = \arctan \frac{\sqrt{(x_d - x_s)^2 + (y_d - y_s)^2}}{h_s}. \quad (1)$$

- (3) Walking intention deviation parameter dev: the rate of change of the supporting force f_d of the arms on the pressure plates is f'_d . When the pressure plates are supported with both arms, the rate of change of the supporting force f'_d will alternate irregularly with the intensity of movement. When the user is walking slowly, the center point P_s of both feet will fluctuate

in a small range near P_d . With increase in walking speed, the linear distance will also increase, and when a fall occurs, the linear distance will rise sharply. This value is taken as the gain of the rate of change of the supporting force, so as to effectively enlarge the deviation parameter of walking intention.

$$\text{Dev} = f'_d \cdot \sqrt{(x_d^2 - x_s^2) + (y_d^2 - y_s^2)}. \quad (2)$$

- (4) The fluctuation frequency f of walking intention deviation parameters: this is the sum of all frequencies of intention deviation within a certain range e in a period of time t . As shown in Figure 6. If the frequency is within this range, it means that the position of both feet still has a large deviation range from the intended direction. From this, we know that both legs did not swing in time and failed to follow the intention of walking direction.

According to the assisted walking experiment, the deviation between the extension line of the actual position and

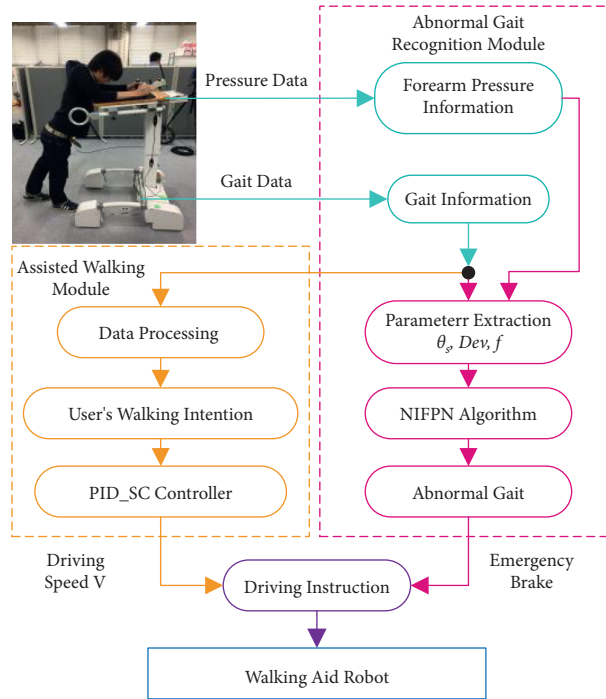


FIGURE 4: The system block diagram.

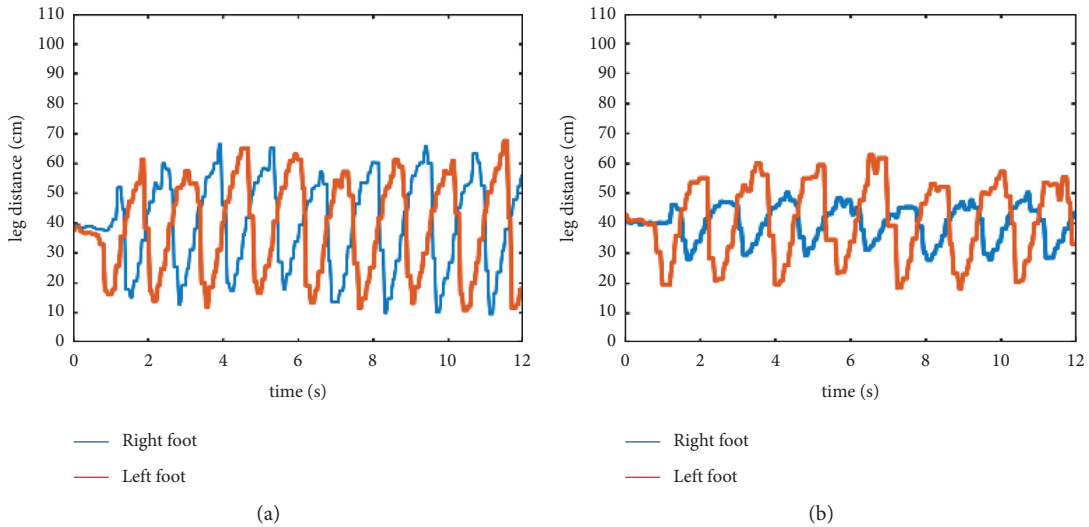


FIGURE 5: Gait information. (a) Normal walking. (b) Restrained walking.

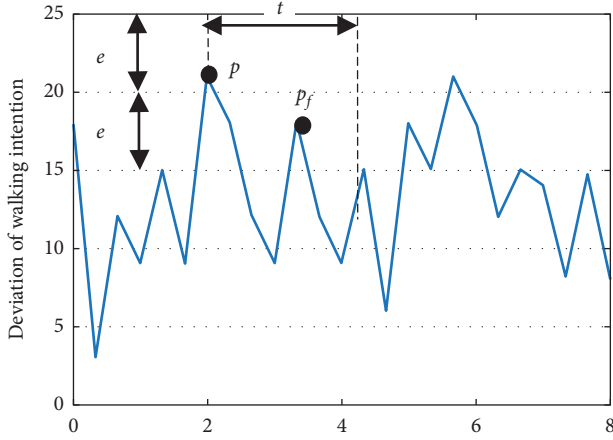
the direction intention is determined by the deviation amount Dev , and the inclination degree of the user's body is determined by the z -axis acceleration value. The frequency f reflects whether the user's gait can stably follow the intended direction in a given period of time.

3.3. Node-Iterative Fuzzy Petri Net Algorithm. Petri net has been widely used in fault diagnosis and other fields. It can effectively describe the dynamic process of abnormal phenomena and has the characteristic advantages of structured expression, quick inference, quick search and mathematical adaptation of diagnosis. The abnormal gait recognition

process of a user while walking is a typical example of a dynamic process.

3.3.1. Fuzzy Petri Net (FPN). FPN is formed from the extension of the basic Petri net idea in [35, 36]. Each library of FPN is assigned a real value on $[0,1]$ as its identification value and each transition is given a definite factor to represent the probability of transition occurrence. The input and output functions are also specified. Here FPN is defined with nine tuples:

$$FPN = \{P, T, D, I, O, \alpha, \beta, Th, U\}. \quad (3)$$


 FIGURE 6: The definition of frequency parameter f .

Here, $P = \{p_1, p_2, \dots, p_m\}$ is a finite set of repository nodes; $T = \{t_1, t_2, \dots, t_m\}$ is a finite set of transition nodes; $D = \{d_1, d_2, \dots, d_m\}$ is a finite set of propositions, and $|P| = |D|$, $P \cap T \cap D = \emptyset$; $I: P \rightarrow T$ is the input matrix; reflects the mapping from library to transition. $I = \{\delta_{ij}\}$, δ_{ij} is a logical quantity, $\delta_{ij} \in \{0, 1\}$, When P_i is the input of T_j (that is, there is a directional arc from P_i to T_j), $\delta_{ij} = 1$; otherwise $\delta_{ij} = 0$, where $i = 1, 2 \dots n, j = 1, 2 \dots m$; $O: T \rightarrow P$, is an output matrix, $O = \{\gamma_{ij}\}$, γ_{ij} is a logical quantity, $\gamma_{ij} \in \{0, 1\}$, when P_i is the output of T_j (that is, there is a directional arc from P_i to T_j), $\gamma_{ij} = 1$; otherwise $\gamma_{ij} = 0$, where $i = 1, 2 \dots n, j = 1, 2 \dots m$; $\alpha: P \rightarrow [0, 1]$, indicates the confidence of the proposition corresponding to the library; $\beta: P \rightarrow D$, is a mapping; reflects the corresponding relationship between the nodes of the library and the proposition; If $\alpha(p_i) = y_i$, $y_i \in [0, 1]$, and $\beta(p_i) = d_i$, the confidence of proposition d_i is y_i . $Th: Th \rightarrow [0, 1]$, defines the domain value λ_i for transition node t_i ($t_i \in T$), $Th = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$. U : rule confidence (CF) matrix, $U = \text{diag}(\mu_1, \dots, \mu_m)$, μ_j is the confidence of rule T_j , $\mu_j \in [0, 1]$.

FPN is a rule-based system, and its rules can be expressed with the corresponding FPN models. In the reasoning of abnormal gait during detection, the rules follow a MISO (multiple-input-single-output) FPN model, as shown in the formula: R_i : IF p_1 OR p_2 OR \dots OR p_m ; THEN p_z (CF = $\mu_1, \mu_2 \dots \mu_m$), because of the possible individual differences between the fuzzy base rule system established for a particular user group and a single patient with mobility difficulties, reasoning accuracy of our algorithm could be reduced. Therefore, it is necessary to modify the base nodes and transition nodes in the original FPN model with nodes that are individual user centered. Please refer to next section for details of the specific node update methods. The corresponding node iterative FPN model is shown in Figure 7.

Among them, the confidence of proposition $p_1, p_2 \dots p_n$ is $\alpha(p_1), \alpha(p_2) \dots \alpha(p_n)$. In the fuzzy reasoning method, the fuzzy product rule is adopted, which describes the fuzzy relationship between the antecedent and result. R is a fuzzy set rule base, $R = \{R_1, R_2, \dots, R_n\}$, i order of the fuzzy rule is R_i .

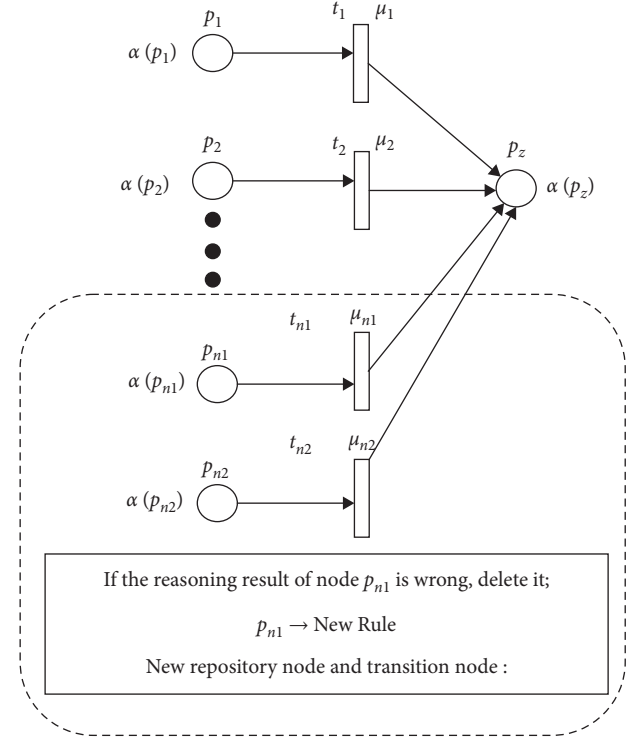


FIGURE 7: The NIFPN model.

3.3.2. Reasoning Flow of the NIFPN. The system input $\text{Mem}(p_i) \forall p_i \in \text{IP}$, IP is set by the input library while the system output is $\text{Mem}(p_i) \forall p_i \in \text{OP}$, and OP is set by the output library. Calculation steps involved in node iterative FPN reasoning of abnormal gait are as follows:

Step 1. Initialization: fuzzy set is defined by membership degree where the initial labeling function is

$$\begin{aligned} M(p_i) &= 0, \text{ if } p_i \notin \text{IP}, \\ M(p_i) &= \text{the number of data tokens, if } p_i \in \text{IP}. \end{aligned} \quad (4)$$

Step 2. Calculate the fuzzy relation matrix, i.e., $\forall t_j \in T$, $V(t_j) = W_a \times W_c = (w_{a1}, w_{a2}, \dots, w_{am})^T \wedge (w_{c1}, w_{c2}, \dots, w_{cm})$, $V(t_j)$ is the fuzzy relation matrix between antecedent and result in a given time t_j , $W_a = \{w_{a1}, w_{a2}, \dots, w_{am}\}$ is the weight fuzzy set of the antecedent while $W_c = \{w_{c1}, w_{c2}, \dots, w_{cm}\}$ is the weight fuzzy set of the result. Each element in a fuzzy set is represented by a fuzzy weight interval.

Step 3. Input the data for detection $W_{a\text{-input}}$.

Step 4. Initiate transition, i.e., calculate

$$\begin{aligned} t_j &\in \frac{T}{\sqrt{P_k}}, \\ W'_a &= W_{a\text{-input}}, \\ W'_c &= W'_a * \circ V(t_j) \text{ or } W'_a \circ V(t_j). \end{aligned} \quad (5)$$

Step 5. Output: for the output variable O , its associated membership function is $W'_c = \{w'_{ci}\} = \forall w'_c, i = 1, 2, \dots, I$; W'_c is the system's output value.

Step 6. When the transition initiation conditions are met, return to Step 4, i.e., meet the following requirements:

$$\exists t_j \in \frac{T}{M(p_i)} = 1, \quad \forall p_j \in I(t_j). \quad (6)$$

Step 7. Calculate the real operation value by using the maximum defuzzification method.

Step 8. If the reasoning result is wrong, delete the original node and regenerate it based on collected data and current state.

3.4. PID_SC Controller. Using the direct distance controller results to a rough, unstable and unsafe movement in the robot [37]. This is caused by an intermittent or discrete motion of the robot in the v_{SC} and y relative position axes. The PID_SC controller proposed serves the purpose of making the motion of robot controller more compliant. As shown in Figure 8, two-dimensional cartesian coordinate system is constructed based on mutually perpendicular multichannel proximity sensors to improve the human-robot interaction process. The geometric center of the robot is $P_{jc}(x_c, y_c)$, the coordinates of the user's left tibia is $P_l(x_l, y_l)$, right tibia is $P_r(x_r, y_r)$, P_p is the next gait position. According to the line from point P_l and P_r , and its midpoint P_{bc} , we defined the body's center of gravity as $P_{bc}(x_b, y_b)$. During assisted walking, we expect P_{jc} and P_{bc} to stay overlapped with each other, that is, the center of the user's body is always near the geometric center of the robot, so as to avoid collision, drag-to gait or overall torso tilt.

First, the PID controller enables the robot to calculate the difference between P_{jc} and P_{bc} . In the process of moving forward or backward, the position error with respect to $P_{bc}(x_b, y_b)$ is represented as e_x, e_y , and e_{yp} (with directions x_j and y_j as reference). They are defined as $e_x = x_j - x_b$ and $e_y = y_j - y_b$ respectively. In order to minimize the error, the controller is designed as

$$\begin{cases} \dot{x}_b = k_{P,x}e_x + k_{I,x} \int e_x dt + k_{D,x}\dot{e}_x, \\ \dot{y}_b = k_{P,y}e_y + k_{I,y} \int e_y dt + k_{D,y}\dot{e}_y. \end{cases} \quad (7)$$

where \dot{x}_b and \dot{y}_b are the input velocity of the system and k_p , k_I , and k_d represent the proportional gain, integral gain and differential gain, respectively. Although the PID is adopted for movement control in this paper, it is yet necessary to input the gait speed of the user to ensure compliance in motion. Reasons being that relative position error changes continuously due to the influence of continuous and unequal gait of the user. That is, during human-robot interaction, geometric relative positions are directly affected by all

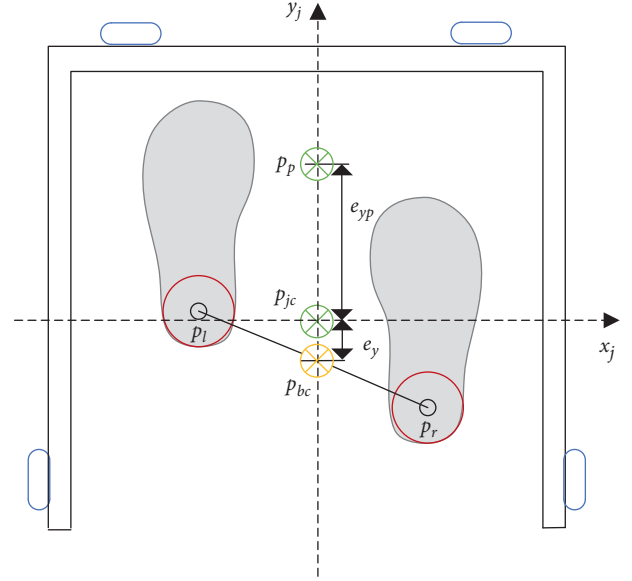


FIGURE 8: Motion information detection.

intermittent motions. Hence, we propose the addition of gait speed compensation in the originally established controller.

Gate speed compensation is mathematically expressed as follows: the user moves the right leg from the initial position to the next position at a distance of d_{sl} and time t , throughout which the left leg is fixed. The rule of thumb is that the step is about twice the displacement of P_{bc} , that is, the absolute velocity $v_a = d_{sl}/1.5t$. The estimated position is $P_p(x_{pc}, y_{pc})$, and then the error in the y direction is $e_{yp} + e_y$. In the right leg movement, the estimated speed v_p is

$$v_p = \frac{e_{yp} + e_y}{t}. \quad (8)$$

Gait speed compensation v_{SC} is calculated by combining the estimated gait speed v_p and the absolute speed v_a expressed as

$$v_{SC} = \frac{d_{sl} + 3e_{yp} + 3e_y}{6t}. \quad (9)$$

For forward or backward movement, the PID_SC controller output is expressed as

$$\dot{y}_b = k_{P,y}e_y + k_{I,y} \int e_y dt + k_{D,y}\dot{e}_y + v_{SC}. \quad (10)$$

4. Experiment and Analysis

The effectiveness of the proposed method is investigated by carrying out comprehensive experiments. 30 subjects with limited mobility (15 males and 15 females) were invited to carry out abnormal gait recognition experiments and assisted walking experiment. Then, the proposed method was tested in smart house to evaluate the safety and comfort of users. Finally, a comparative analysis was carried out to prove the superiority of the proposed method.

TABLE 1: Parameters of the membership function.

Input parameter	Low			Middle			High	
Dev	12	16	14	20	26	22	28	
z	5	16	11	18	24	21	27	
f	1	3	2	4	6	5	6	

4.1. Abnormal Gait FPN Model. According to the walking intention recognition and parameter extraction, the membership function is defined by the deviation Dev, z -axis acceleration, and frequency f . The function is divided into three states: high, medium, and low. To meet actual requirements high, medium, and low membership function parameters are divided according to the average deviation, as shown in Table 1, and their corresponding membership functions are shown in Figure 9.

In this paper, the logical relationships among the gait information, position and movement direction deviations, and parameters with self-adjustment capabilities are simplified based on FPN and represented with “library” and “transition” nodes which are connected by directional arcs. Experimental analysis shows that the value of Dev is usually between 8 and 30, the body inclination angle z decreases to values between 3 and 28, and the frequency f is between 1 and 10. When walking occurs slowly, Dev is usually between 10 and 16, inclination angle z , between 5 and 16, and frequency f , and between 5 and 6. During a fall, Dev usually has values of range 22 ~ 28, the torso inclination angle z will be between 21 and 27 and the frequency f decrease to a range of 1 ~ 3. Because the member function is required to be between 0 and 1, the three input parameters are normalized to the range of 0 to 1. α, β , and γ represent Dev, z -axis acceleration and frequency parameter f , respectively. H, M , and L represent the membership functions of “high,” “medium,” and “low.”

The fuzzification process is defined as follows: three fuzzy rules are formulated to correspond with three fuzzy results: normal, fast and abnormal walking gait. Next, we setup and configure input language variables, deviation Dev, torso inclination angle z and frequency parameter f . The different language variables are defined as high, medium and low. Fuzzy rules are

$$\begin{aligned} R_1: & \text{ if Dev is } L \text{ and } z \text{ is } L \text{ and } f \text{ is } H \text{ Then } D \text{ is } NA, \\ R_3: & \text{ if Dev is } H \text{ and } z \text{ is } H \text{ and } f \text{ is } L \text{ Then } D \text{ is } F. \end{aligned} \quad (11)$$

FPN transformation result based on the above fuzzy rules is shown in Figure 10.

4.2. Case Analysis of Abnormal Gait. For illustrative purposes, we use the above mention FPN calculation steps to analyze abnormal gaits. The matrix of Dev, z and F parameters in the reasoning process is Dev, z , and F , respectively.

Step 9. Set the fuzzy set according to the experiment:

$$\begin{aligned} Dev_L &= \frac{0.35}{dev_{ll}} + \frac{0}{dev_{lm}} + \frac{0}{dev_{lh}}, \\ Z_L &= \frac{0.27}{z_{ll}} + \frac{0}{z_{lm}} + \frac{0}{z_{lh}}, \\ F_M &= \frac{0.42}{f_{ml}} + \frac{0}{f_{mm}} + \frac{0}{f_{mh}}, \\ Dev_M &= \frac{0}{dev_{ml}} + \frac{0.58}{dev_{mm}} + \frac{0}{dev_{mh}}, \\ Z_M &= \frac{0}{z_{ml}} + \frac{0.49}{z_{mm}} + \frac{0}{z_{mh}}, \\ F_H &= \frac{0}{f_{hl}} + \frac{0.60}{f_{hm}} + \frac{0}{f_{hh}}, \\ Dev_H &= \frac{0}{dev_{hl}} + \frac{0}{dev_{hm}} + \frac{0.81}{dev_{hh}}, \\ Z_H &= \frac{0}{z_{hl}} + \frac{0}{z_{hm}} + \frac{0.71}{z_{hh}}, \\ F_L &= \frac{0}{f_{lh}} + \frac{0}{f_{lm}} + \frac{0.76}{f_{ll}}, \\ Status &= \frac{0.35}{s_l} + \frac{0.5}{s_m} + \frac{0.75}{s_h}. \end{aligned} \quad (12)$$

Step 10. Calculate the Cartesian product of the antecedent and the result to obtain the fuzzy relation matrix;

$$\begin{aligned} P_1 &= Dev_L \times Z_L \times F_M = \left((0.35 \ 0 \ 0)^T \wedge (0.27 \ 0 \ 0) \right)^T \wedge (0.42 \ 0 \ 0) \\ &= \begin{bmatrix} 0.27 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}^T \wedge (0.42 \ 0 \ 0) = \begin{bmatrix} 0.27 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \end{aligned}$$

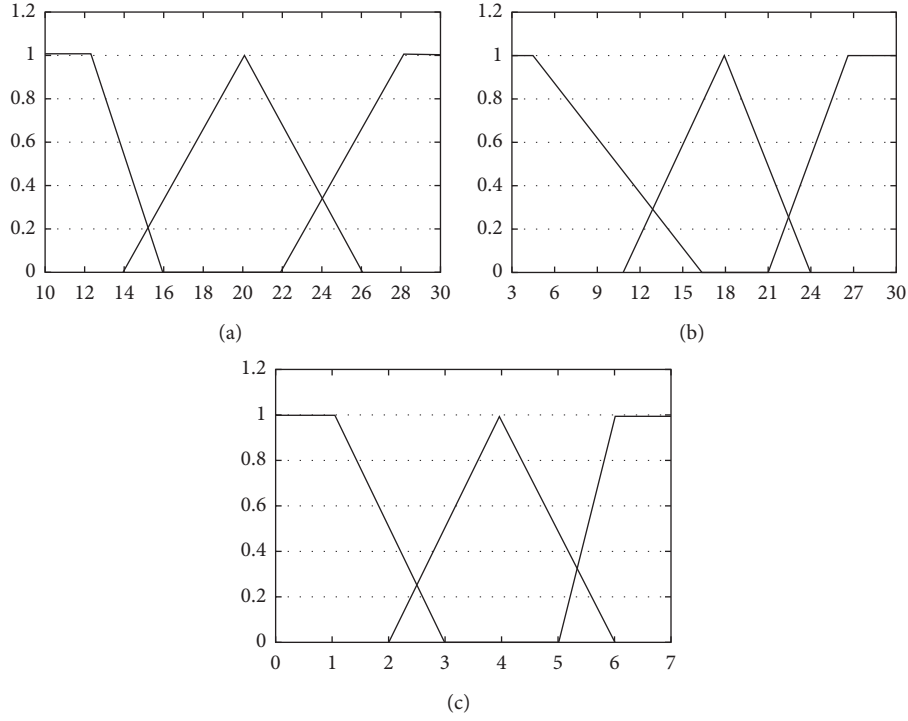


FIGURE 9: Membership function for input parameter. (a) Membership function for Dev. (b) Membership function for z. (c) Membership function for f.

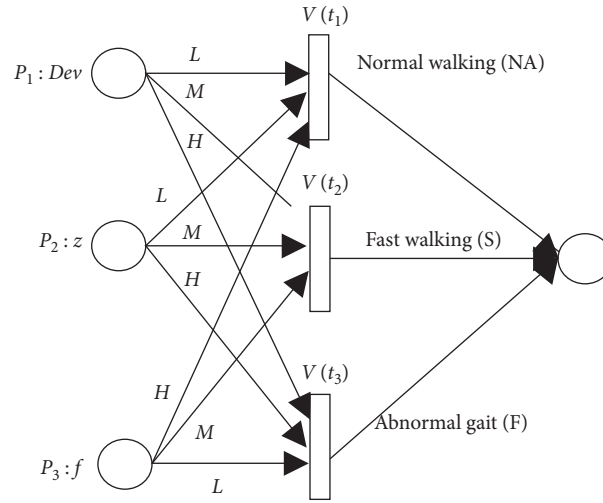


FIGURE 10: The fuzzy Petri model of abnormal gait detection.

$$\begin{aligned}
 P_2 &= Dev_M \times Z_M \times F_H = \left((0 \ 0.58 \ 0)^T \wedge (0 \ 0.49 \ 0) \right)^T \wedge (0 \ 0.6 \ 0) \\
 &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.49 & 0 \\ 0 & 0 & 0 \end{bmatrix}^T \wedge (0 \ 0.6 \ 0) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.49 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\
 P_3 &= Dev_H \times Z_H \times F_L = \left((0 \ 0 \ 0.81)^T \wedge (0 \ 0 \ 0.71) \right)^T \wedge (0 \ 0 \ 0.76) \\
 &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.71 \end{bmatrix}^T \wedge (0 \ 0 \ 0.76) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.71 \end{bmatrix}.
 \end{aligned}$$

$\forall t_j \in T$, T is the transition set. Calculate fuzzy relation matrix $V(t_j)$, which is the matrix of fuzzy antecedent and rule t_j , as follows:

$$\begin{aligned} V(t_1) &= \begin{bmatrix} 0.27 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in P_1 \times \text{Status} \times s_p, \\ V(t_2) &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.49 & 0 \\ 0 & 0 & 0 \end{bmatrix} \in P_2 \times \text{Status} \times s_m, \\ V(t_3) &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.71 \end{bmatrix} \in P_3 \times \text{Status} \times s_h. \end{aligned} \quad (14)$$

Step 11. Enter the set to be detected

$$\begin{aligned} \text{Dev}' &= \frac{0}{\text{dev}_l} + \frac{0.02}{\text{dev}_m} + \frac{0.65}{\text{dev}_h}, \\ Z' &= \frac{0}{z_l} + \frac{0}{z_m} + \frac{0.63}{z_h}, \\ F' &= \frac{0}{f_m} + \frac{0}{f_h} + \frac{1}{f_l}. \end{aligned} \quad (15)$$

Step 12. Transition trigger changes

$$\begin{aligned} S'_1 &= \text{Dev}' \circ V(t_1) \circ Z' \circ F' \\ &= [0 \ 0.02 \ 0.65] \circ \begin{bmatrix} 0.27 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \circ [0 \ 0 \ 0.63] \circ [0 \ 0 \ 1] \\ S'_2 &= \text{Dev}' \circ V(t_2) \circ Z' \circ F' \\ &= [0 \ 0.02 \ 0.65] \circ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.49 & 0 \\ 0 & 0 & 0 \end{bmatrix} \circ [0 \ 0 \ 0.63] \circ [0 \ 0 \ 1] \\ S'_3 &= \text{Dev}' \circ V(t_3) \circ Z' \circ F' \\ &= [0 \ 0.02 \ 0.65] \circ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.71 \end{bmatrix} \circ [0 \ 0 \ 0.63] \circ [0 \ 0 \ 1] \end{aligned} \quad (16)$$

Step 13. The results of fuzzy reasoning are as follows:

$$D = S'_1 \cup S'_2 \cup S'_3 = \frac{0}{d_1} + \frac{0.02}{d_2} + \frac{0.63}{d_3}. \quad (17)$$

Step 14. The actual selected value is determined by using the maximum defuzzification method. Because d_3 has the maximum membership degree, the case is regarded as a fall.

Step 15. Collect the abnormal gait of a subject whose right knee is fixed when he or she falls. The system's input data set to be analyzed for detection is

$$\begin{aligned} \text{Dev}' &= \frac{0}{\text{dev}_l} + \frac{0.5}{\text{dev}_m} + \frac{0.17}{\text{dev}_h}, \\ Z' &= \frac{0}{z_l} + \frac{0.17}{z_m} + \frac{0.33}{z_h}, \\ F' &= \frac{0}{f_m} + \frac{0}{f_h} + \frac{1}{f_l}. \end{aligned} \quad (18)$$

According to the same reasoning method, the result of the system is

$$D = S'_1 \cup S'_2 \cup S'_3 = \frac{0}{d_1} + \frac{0.17}{d_2} + \frac{0.17}{d_3}. \quad (19)$$

That is to say, fall and fast walking inference probability obtained by the system is obviously inconsistent with the falling situations of the subject. The reason is that the Dev value of 22.98 of an abnormal gait is more similar to the fast gait of healthy subjects. Because the knee is limited and the gait state is different from that of healthy people, the inference accuracy will be reduced by using the same inference parameters. Therefore, it is necessary to adaptively update the transition trigger parameters in FPN. The updating method is shown in Figure 11. The membership degree of 22.98 in $de v_h$ is set as 1, and the membership degree of $de v_m$ in 22 is set as 0, so that the membership degree of $de v_m$ moves to the left as a whole with 18 as the center. That is, the membership degree corresponding to different values is replanned, thus updating the nodes confidence. The reasoning result after node update is

$$D = S'_1 \cup S'_2 \cup S'_3 = \frac{0}{d_1} + \frac{0.17}{d_2} + \frac{0.33}{d_3}. \quad (20)$$

That is, the actual value determined by the maximum defuzzification method is of an actual fall situation.

4.3. Abnormal Gait Recognition Experiment. Ten subjects wore devices at their knees to imitate the daily gait of patients with lower limb impairment to conducted indoor normal speed (0.27 m/s), fast speed (0.58 m/s), and abnormal gait walking experiments. The values $e = 4$ and $t = 3$ were constant. Figure 12 shows the deviation parameters of walking intention of the subject when using the robot to perform experiment in the above listed conditions. When a subject is walking normally, the deviation parameters of the walking intention were consistent. As shown in Figures 12(b) and 12(e), the five peaks of Dev are 19.32, 18.79, 17.53, 15.26, and 17.88, which occur at 2.6 s, 3.2 s, 3.9 s, 4.6 s, and 5.2 s respectively. Therefore, the maximum value, 19.32, of Dev, and the maximum value, 13, of the torso inclination angle Z within 3 seconds period, both serves as the system input. Hence, the system inputs are $\text{Dev} = 19.32$, $z = 13$, and $F = 5$, the reasoning results of the system are $S_1 = 0$, $S_2 = 0.45$, and $S_3 = 0$. Final results obtained from maximum defuzzification method indicates that it was a fast

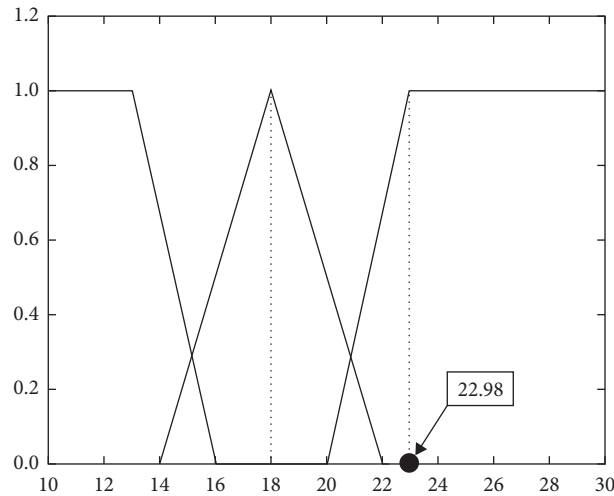


FIGURE 11: Membership function for input parameter.

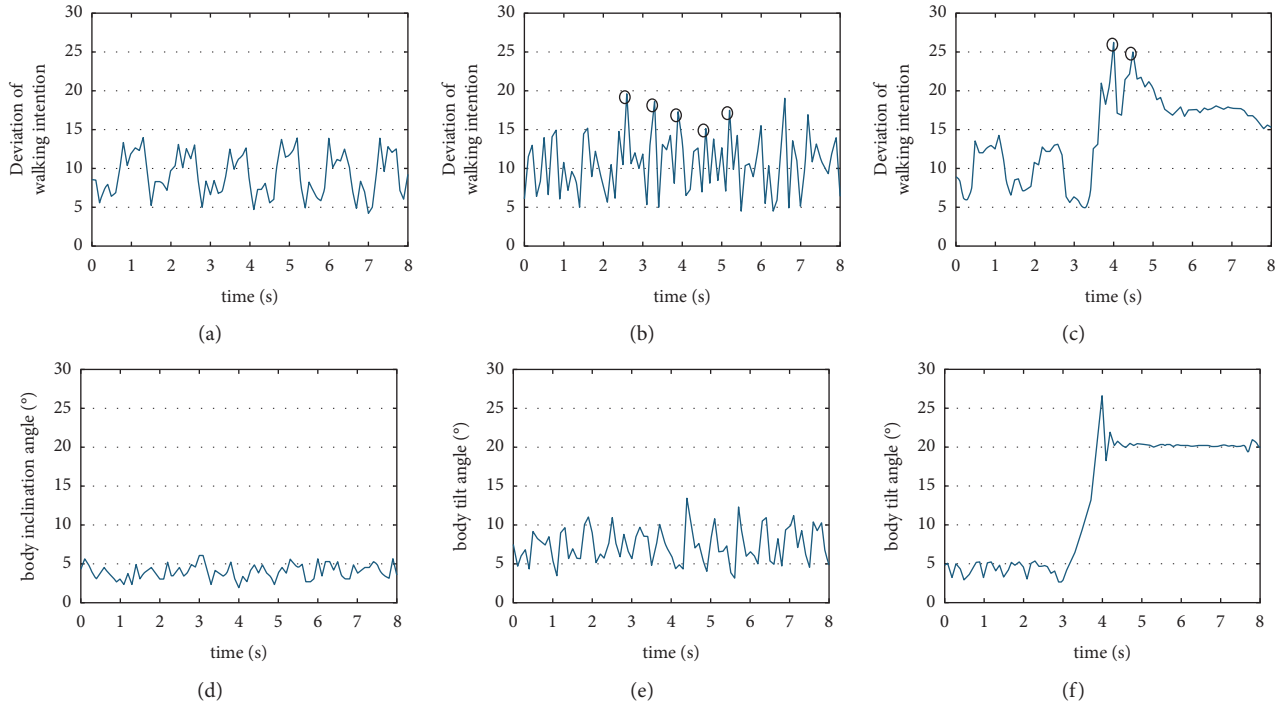


FIGURE 12: Normal walk, fast walk and fall of subject. (a) Dev of normal walk. (b) Dev of fast walk. (c) Dev of fall. (d) Z of normal walk. (e) Z of fast walk. (f) Z of fall.

speed walking case. Similarly, as shown in Figures 12(c) and 12(f), when a drag-to-gait occurred, the two peaks of Dev are 26.91 and 25.01, which happen at time 3.9 s and 4.3 s, respectively. Again, the maximum value of Dev which is 26.91 and the maximum value of body torso inclination angle Z within 1 second period which is 27, both serve as the system's input. With $Dev = 26.91$, $z = 27$, and $F = 2$, the reasoning result of the system is $S_1 = 0$, $S_2 = 0.2$, and $S_3 = 0.43$. Final results obtain from maximum defuzzification method indicates that this is an abnormal gait walking case.

Figure 13 shows the abnormal gait recognition experimental data result process of a subject with a fixed right knee. The experiment begins when the subject is at a standstill position. From t_0 to t_1 , the distance between both legs changes abruptly. Also, the intention direction deviation degree changes greatly because the subject is constantly adjusting his gait to balance the upper body from the start to the end of the test process with the robot. Hence, data from the first 2 seconds of the experiment were not passed to the system as input until t_2 when the subject begins to walk normally. The drag-to-gait occurring within t_2 to t_3 was

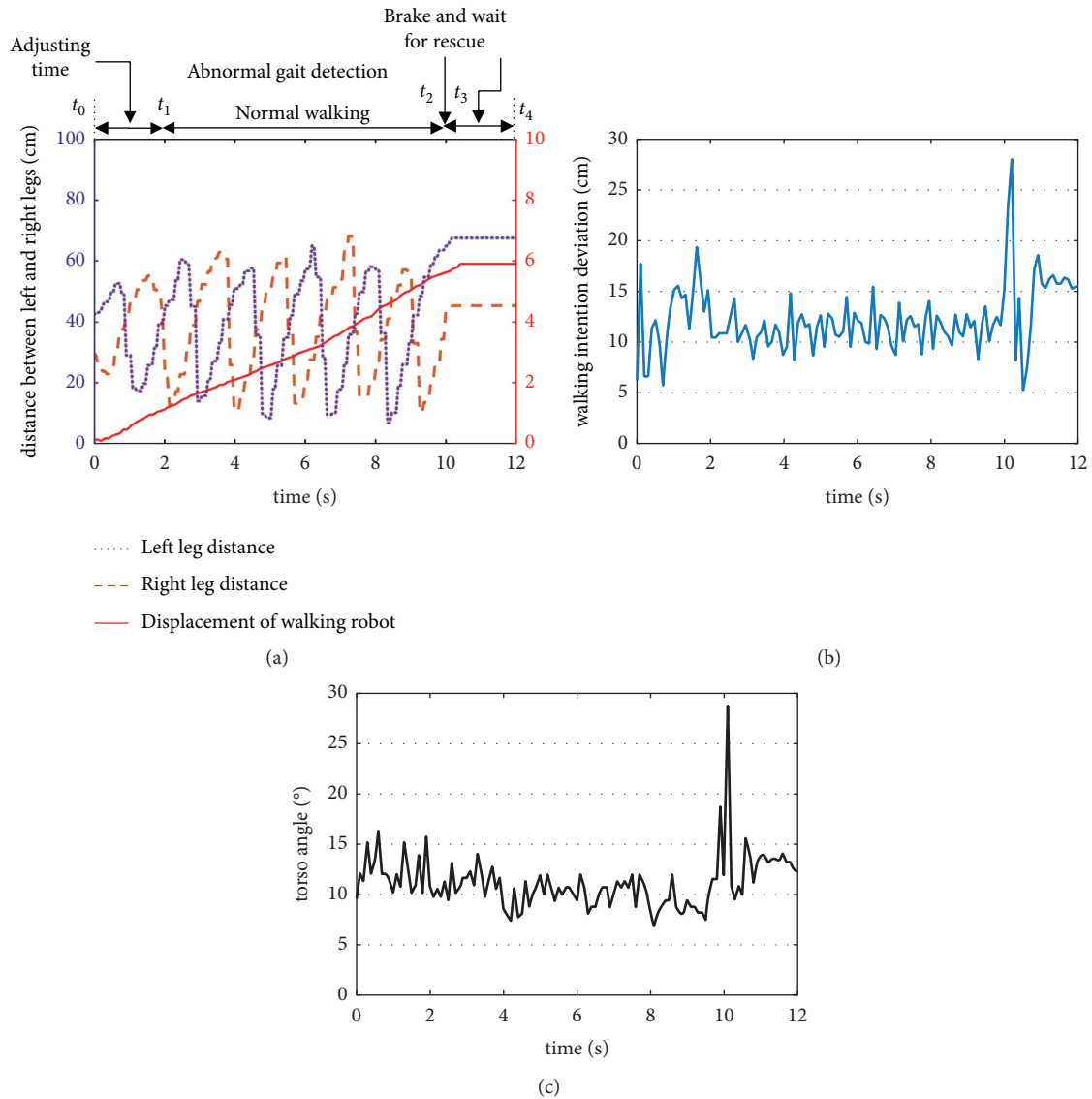


FIGURE 13: Parameter's variation of normal walk of health people. (a) Gait information and robot displacement. (b) Deviation parameter Dev of walking intention. (c) Slope angle Z of body.

recognized by the system. Here, the peak value of Dev is 28.93, time at 10.1 s. So, the input of the system are $Dev = 28.03$, $z = 29$, and $F = 1$. The system's reasoning are $S_1 = 0$, $S_2 = 0$, and $S_3 = 0.78$. After maximum defuzzification, reasoning results indicate that it is an abnormal walking gait. Thus, at t_3 , the robot brakes urgently and rings a rescue alarm while the subject waits for rescue.

We comprehensively compare the NIFPN algorithm with SVM algorithm[38, 39]. Taking the falls and daily routines of general users into consideration, the abnormal gait test was conducted. To ensure the safety of the whole experiment, an elastic bandage is fixed between the subject and the robot to ensure that all the subject will not fall completely. The abnormal gait are categorized into forward falls, backward falls, vertical falls, and sideway falls, and drag-to gaits. We compare the two algorithms using the accuracy rate and misidentification rate. The conducted experiments and results are summarized in Table 2.

The accuracy percentage of abnormal gait of NIFPN is up to 91.2%, and the recognition rate of drag-to gait is much higher than SVM algorithm particularly. The misidentification percentage of daily routines is 6.27%. The majority of fall misjudgment is more likely to occur with sideway fall due to the relatively lower SVM generated at sideway fall, and therefore more easily to lead to misjudgment. Through the above experiments, the advantages of NIFPN algorithm are summarized as follows:

- (1) The input parameters of the algorithm reflect the relative position information between the user and the robot, and it replaces the experience-based threshold in traditional fall recognition methods. Therefore, this algorithm does not need to store the abnormal behavior gait data of users in advance, which improves the universality of the algorithm.

TABLE 2: The contrast results of the NIFPN and SVM.

Events	Action times		Detected abnormal gait		Detected nonabnormal		Accuracy rate	
	NIFPN	SVM	NIFPN	SVM	NIFPN	SVM	NIFPN	SVM
Forward falling	30		25	25	0	0	100%	100%
Backward falling	30		29	25	1	0	96.7%	100%
Sideway falling	30		25	27	5	3	83.3%	90%
Vertical falling	30		27	29	3	1	90%	96.7%
Walking	30		0	0	7	1	100%	100%
Drag to gait	30		26	12	4	18	86.7%	40%

- (2) The algorithm introduces node iteration algorithm, which can effectively solve the differences of user with different gait habit. The recognition rate is improved by adaptive updating of nodes.
- (3) The algorithm can effectively recognize the drag-to gait, which is more suitable for the real use scenario.

4.4. Assisted Walking Experiment. To better serve the elderly and disabled, the laboratory setup a smart house environment [40] with a variety of welfare robots, such as gait rehabilitation robot, walking aid robot, intelligent wheelchair robot, transport robot, and excretion support robot as shown in Figure 14. It consists of 3 areas: recreation area, living area, and rehabilitation area. The control methods proposed in this paper are integrated in these robots.

To verify the effectiveness of the proposed noncontact compliant control method, 4 subjects conducted a multi-directional trajectory tracking experiment. The subjects walked in 8 directions assisted by the WAR according to preset trajectory. The walking path are made in square and diamond shapes with side length of 2 meters each. At the same time, some areas were marked with yellow markers, indicating that subjects should slow down when crossing these areas.

Figure 15 comparatively shows the path results of the four subjects while using WAR with respect to the target walking path. Although there are slight differences between the walking trajectory of the subjects and the preset trajectory, experimental results show that the method can accurately identify the subjects' walking intention direction and thus is able to satisfy rehabilitative needs.

Figure 16 shows the "slow-fast-slow" walking gait results of two subjects along the preset path. Although the subjects have individual differences, the robot can closely follow their walking gait.

To verify the superiority of the proposed method, the contrast experiments between the PID_SC controller, traditional PID controller and DDC (direct detection controller [37]) were implemented, as shown in Figure 17. Before the experiment, subjects were first allowed to use the robot for half an hour to ensure that they were fully familiar with the WAR operation mode. To ensure safety the maximum driving speed of WAR was set to 1.1 m/s. Then, all subjects were asked to walk a distance of 20 meters along the preset path using three compared control methods, respectively. The gait data and robot displacement data were recorded.

The comparative data of all the subjects with three control methods are shown in Figure 18. The displacement differences of all healthy subjects by PID_SC, PID, and DDC are 2.63 cm, 4.19 cm, and 4.96 cm, respectively. The displacement differences of subjects with limited mobility are 3.34 cm, 4.72 cm, and 5.63 cm, respectively.

The experimental results show that the displacement error of the compliant control method proposed in this paper is smaller than that of the traditional PID controller and direct distance controller. Faced with subjects with different motion capabilities, the proposed controller can control the WAR to produce movement closely corresponding to the user's walking gait. It is observed that the gait length and frequency of the healthy subjects can maintain a high consistency (almost uniform in speed), while the gait length and frequency of the subjects with limited mobility have more observable changes (long or short gait, and frequency fluctuation). Therefore, the overall displacement error of the subjects with limited mobility is slightly larger than that of the healthy subjects. Compared with the traditional control method, the controller introduces human gait speed compensation to reduce the relative displacement error between the robot and the user. Meanwhile, it also reduces the motion jam phenomenon and makes the motion process more flexible.

4.5. Comparison Analysis Experiment. To test the effectiveness of walking intention-based compliant control in gait rehabilitation, Tekscan Walkway footpath detection system was used to conduct assisted walking gait phase analysis experiment on test subjects [41]. We selected 20 subjects to participate in this experiment. All subjects were informed in advance and they agreed to all the test procedures of the experiment. In common practice gait phase is divided into eight. Since the subject's foot does not bear any pressure during the initial swing phase, the middle swing phase and the final swing phase while walking, these three phases were uniformly referred to as the swing phase for convenience in subsequent work. Table 3 shows the data results of the subjects when performing compliant assisted walking compared to the passive assisted walking with the robot.

The data in the table are expressed in mean \pm standard deviation. Perform difference analysis on the data in the table. $*P < 0.05$ indicates significant difference, and $**P < 0.01$ indicates extremely significant difference. For passive assisted walking, the movement path and speed of

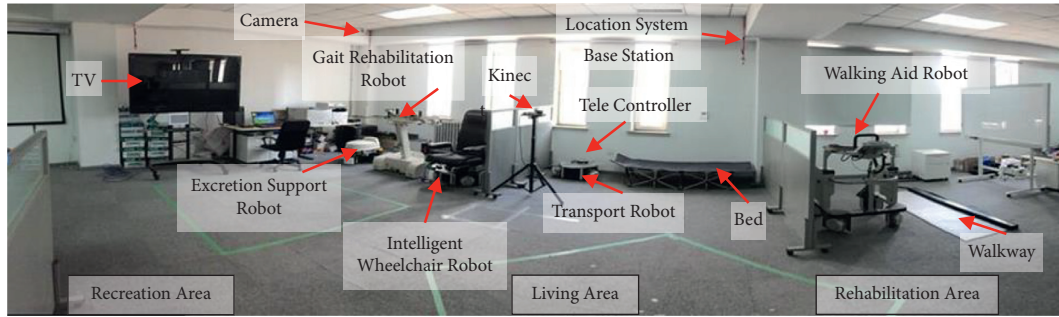


FIGURE 14: Smart house scenario based on multiply welfare robots.

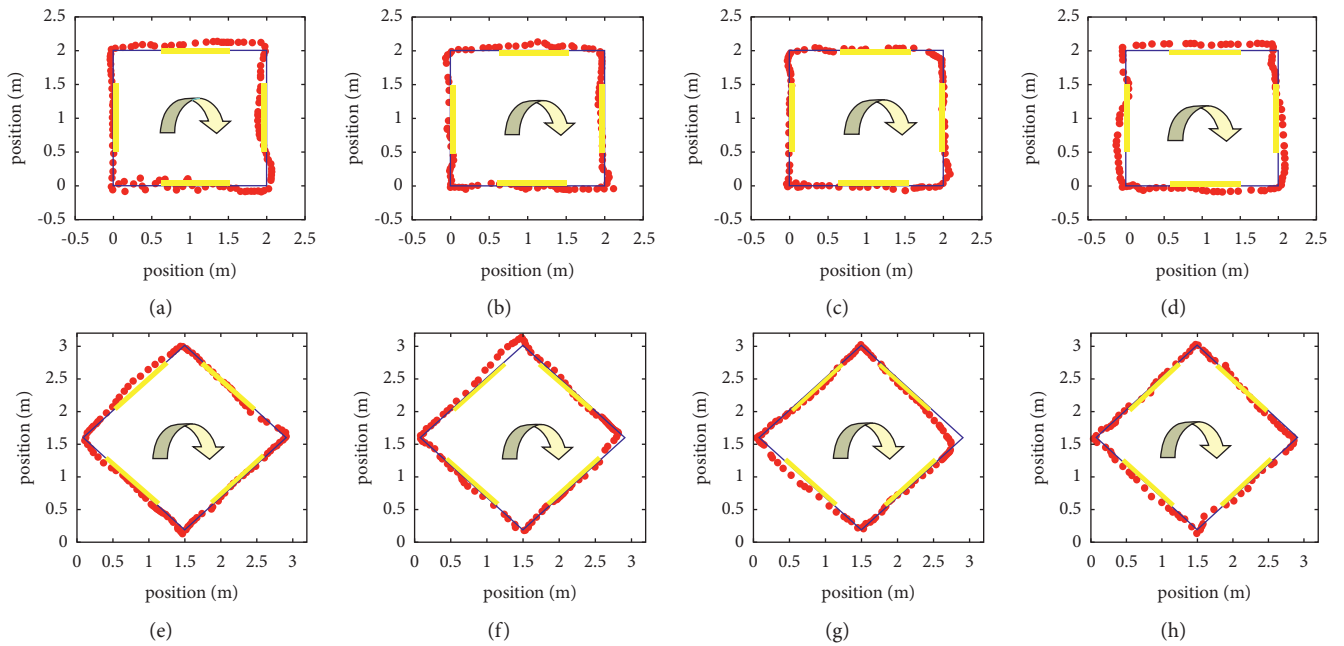


FIGURE 15: Preset path following experiment. (a) Square path of subject A. (b) Square path of subject B. (c) Square path of subject C. (d) Square path of subject D. (e) Diamond path of subject A. (f) Diamond path of subject B. (g) Diamond path of subject C. (h) Diamond path of subject D.

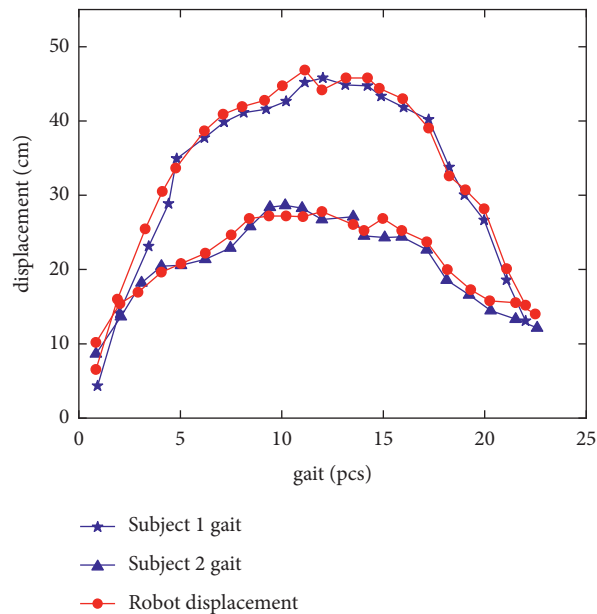


FIGURE 16: Experimental results of gait following.



FIGURE 17: Auxiliary walking experiment.

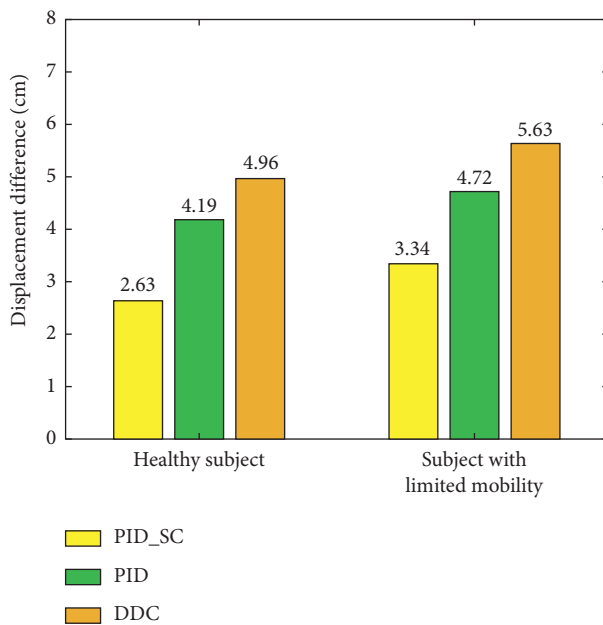


FIGURE 18: The contrast experiments of the PID_SC, PID, and DDC.

the robot are set in advance while the subject follows the robot to complete the rehabilitative walking exercise. The experimental results showed that the single leg support time and double leg support time of the subject were significantly increased when the compliant walking aid was used. Compared with passive assisted walking, the initial leg support time and swing time of the nonaffected leg increased significantly, while other time parameters showed no significant difference. The contact area, pressure and peak pressure of the two assistance methods are shown in Table 4.

The contact area between the midfoot and the full foot of the affected side increased significantly when the subject used robotic assisted compliant walking, but there was no significant difference between the forefoot and the rear foot.

At the same time, there was a significant increase in the subject's midfoot and full foot contact pressure value. The nonaffected midfoot pressure and the peak pressure was higher in the compliant assisted walking than in the passive assisted walking through this index was not obvious on the affected side. These results indicated that subjects can actively walk and master their own gait rhythm during walking rehabilitation which significantly improves the symmetry and stability of the user's gait. Additionally, the improved safety protective measures in human-robot interaction provides psychological guarantee and reduces the mental and physical burden associated in using the robot for passive assisted walking. Overall, subjects show a boost in confidence while using GRR in walking which is necessary for medical rehabilitative recovery [42].

The asymmetrical analysis of all subjects' gaits is shown in Table 5. When subjects are assisted with robot compliance, the asymmetrical index of contact area, standing time, and swing time of both lower limbs were significantly improved compared to that of the passive assistance, given that the asymmetry index of contact pressure and track length has no significant difference. Results shows that robot-assisted gait training meets rehabilitation requirement and can significantly improve the user's gaits symmetry and stability.

Because the gait information detection system developed in this paper can be directly integrated into the mobile chassis and support plate, it can be directly applied to the wheeled walking aid robot with similar structure, and has good universality. The method proposed in this paper was compared with traditional identification methods which require special wearable motion detection device such as pressure sensor and gyroscope, as shown in Table 6.

The approach used in this paper which is based on proximity sensor and pressure sensor basically has the same recognition rate as other abnormal gait recognition methods. Again, it is worth mentioning that this method can recognize all occurring drag-to gaits, and since this method does not require the user to wear any sensor, it increases the user's comfort and convenience. From the subjective point of view of users, this section makes a quantitative investigation on the comfort and acceptability of noncontact interaction methods. Aiming at measuring the robot-induced stress on humans during coexistence, subjective evaluation is usually acquired [43, 44]. During the comprehensive experiment, the comfortable feeling of different interaction methods is evaluated by a questionnaire result from all the subjects, which verifies the effectiveness and comfort of the proposed method. In the one-to-six scale, a higher score means a better comfortable feeling. Table 7 shows the score change of comfortable feeling from the questionnaire survey. The "↑" represents an improvement in comfort, and the "=" and "↓" represent no significant change in comfort or less comfort than the previous methods. From this survey, we can find that most subjects felt more comfortable after an adjustment than traditional wearable methods before. This is because the proposed algorithm is based on the noncontact interaction method. The user can control the robot naturally without the repeated steps of placing the wearable sensor.

TABLE 3: Comparison of gait phase between compliant robotic-assisted walking and passive walking.

Parameter	Affected side		Nonaffected side	
	Compliant assisted	Passive assisted	Compliant assisted	Passive assisted
Support phase	2.11 ± 0.74	1.62 ± 0.52	1.96 ± 0.75	1.57 ± 0.38
Initial leg support	0.46 ± 0.56	0.37 ± 0.37	0.74 ± 0.84**	0.24 ± 0.36
Single leg support	1.21 ± 0.45**	0.57 ± 0.65	0.97 ± 0.51	0.86 ± 0.58
End leg support	0.75 ± 0.46**	0.22 ± 0.46	0.46 ± 0.75	0.34 ± 0.46
Swing phase	0.89 ± 0.37	0.79 ± 0.47	1.31 ± 0.45**	0.71 ± 0.15

TABLE 4: Comparison of contact area and pressure between compliant robotic-assisted walking and passive walking.

Parameter	Position	Affected side		Nonaffected side	
		Compliant assisted	Passive assisted	Compliant assisted	Passive assisted
Contact area (cm ²)	Forefoot	29.75 ± 14.48	25.74 ± 17.76	32.46 ± 13.25	29.85 ± 14.65
	Mid-foot	46.74 ± 15.37**	31.37 ± 16.84	43.64 ± 13.65	40.75 ± 16.75
	Rear foot	30.44 ± 13.52	26.43 ± 12.64	31.65 ± 17.75	33.01 ± 12.36
	Full foot	106.93 ± 26.74**	83.54 ± 24.16	107.75 ± 23.33	103.61 ± 32.03
Contact pressure (10 ³ kPa)	Forefoot	48.36 ± 39.96	51.25 ± 41.64	87.37 ± 47.76	77.86 ± 81.04
	Mid-foot	94.64 ± 48.64**	59.96 ± 39.63	92.73 ± 58.76*	71.65 ± 74.18
	Rear foot	59.37 ± 36.52	42.65 ± 26.37	89.97 ± 71.27	70.75 ± 75.64
	Full foot	202.37 ± 133.05*	153.86 ± 77.27	270.07 ± 85.74**	220.26 ± 94.72
Peak pressure (10 ³ kPa)		218.53 ± 135.64	172.36 ± 82.46	310.73 ± 128.17**	259.36 ± 127.84

TABLE 5: Comparison of asymmetric index between compliant robotic-assisted walking and passive walking.

Parameter	Compliant assisted	Passive assisted
Support time	0.12 ± 0.47*	-0.46 ± 0.78
Swing time	-0.55 ± 1.19*	0.41 ± 0.69
Contact area	-7.54 ± 19.65*	-27.76 ± 16.35
Contact pressure	-38.75 ± 53.46	-45.54 ± 35.57

TABLE 6: : Comparison of human robot interaction methods.

Parameter	Pressure sensor	Gyroscope	The proposed method
Wearable devices	Special wearable devices	Special wearable devices	No
Wear position	Insole	Limb or trunk	No
Compliant control	⊗	√	√
Abnormal gait	√	√	√
Drag-to gait	⊗	⊗	√
Accuracy	90%	95.83%	91.2%
Misidentification rate	18.18%	0.89%	6.27%

Meanwhile, brain monitoring techniques have the capability to detect and characterize the operator's mental state such as workload, fatigue, or mental stress [45, 46]. It has been applied to assisted driving and assisted rehabilitation training for behavior correction and enhancing the acceptability of human-robot interaction. In this paper, Functional Near-Infrared Spectroscopy (fNIRS) WOT-100, a brain imaging system to perform a continuous measure of the mental state, is introduced to monitor the user's mental fatigue when implementing noncontact interaction method and traditional method, as shown in Figure 19.

The mean and peak values of oxygenated signal are extracted as classification features, and the continuous autonomous assisted behaviors are classified combined with

linear discriminant classifier LDA (linear discriminant analysis). As a method to evaluate mental fatigue, the classification results can directly distinguish the mental states of two different difficulty levels of behavior. The experiments of noncontact interaction method and traditional wearable interaction methods were carried out on 20 subjects for seven times. The classification results are shown in Table 7. The classification results of cerebral blood oxygen parameters of two tasks with different difficulty levels are large, and the difference is obvious. The greater physical exertion, weak action ability and balance ability subjectively cause the psychological load of subjects on risk behaviors such as falls, and increase the fluctuation of cerebral blood oxygen parameters. The noncontact interaction method can

TABLE 7: Comparison of subjective parameters of different human-robot interaction methods.

Subject	Score (traditional Noncontact)	Trend	Classification Results
1	4/6	↑	74.53
2	3/5	↑	81.36
3	5/5	=	63.13
4	3/4	↑	69.64
5	4/6	↑	70.34
6	3/4	↑	75.24
7	1/5	↑	87.37
8	2/6	↑	86.43
9	4/3	↓	59.76
10	4/6	↑	81.35
11	4/6	↑	69.76
12	2/4	↑	73.58
13	4/4	=	55.67
14	2/4	↑	76.53
15	2/6	↑	81.87
16	3/6	↑	72.76
17	2/4	↑	57.34
18	3/4	↑	79.48
19	4/3	↓	82.84
20	5/6	↑	75.78

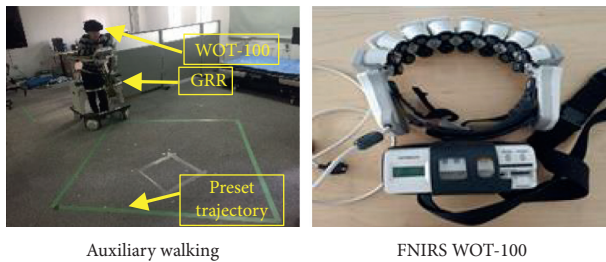


FIGURE 19: Acceptability experiment of human-robot interaction. (a) Auxiliary walking. (b) FNIRS WOT.

significantly reduce the physical consumption and mental fatigue of the subjects.

In general, the advantages of the noncontact interaction method proposed in this paper are summarized as follows:

- (1) The proposed method can ensure the user's direct operation, and avoids the cumbersome steps of repeated wearing and data correction. It enhances the user's comfort and convention.
- (2) For gait rehabilitation training, the proposed method helps users to walk actively and master their gait rhythm during rehabilitation process, which significantly improves the symmetry and stability of the user's gait.
- (3) The noncontact interaction method provides psychological guarantee, and reduces the mental and physical burden associated in using the robot. Overall, subjects show a boost in confidence while interacting with the robot with our proposed method, which is necessary for auxiliary walking and medical rehabilitative recovery.

5. Conclusion

This paper proposes a safe and compliant noncontact interactive approach for the wheeled walking aid robot. First, combined with the mechanical structure of wheeled walking aid robot, an expandable multichannel proximity sensor is designed. These sensors are combined and installed in the robot mobile chassis to recognize human gait information effectively. Secondly, a noncontact abnormal gait recognition approach based on NIFPN algorithm is proposed which identifies asynchronous human-robot movement speed or physical impairment induced falls and drag-to gaits during walking, enabling the robot to brake in emergency situations so as to ensure the safety of the user. Then, a PID_SC controller which integrates gait speed compensation feature is designed to accurately and compliantly follow the user's gaits. Experimental results show that the NIFPN algorithm can accurately identify abnormal gaits of groups with different walking habits, and the recognition rate reaches 91.2%. Moreover, the designed PID_SC controller significantly improves the compliance and stability of the robot during assisted walking. Considering the convenience and comfort that the method offers by not requiring patients to wear sensors that introduce troublesome step of detection point pre-correction, it can be applied to all wheeled walking aid robots with similar structures, and popularized to help the elderly and the disabled in hospitals, home and other places.

In the next step, we will further explore the safety of wheeled walking aid robot. Our purpose is to develop an environmental and gait information-based controller which will encourage safety in small space areas such as homes.

Data Availability

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] L. Peng, Z. G. Hou, C. Wang, L. C. Luo, and W. Q. Wang, "Physical interaction methods for rehabilitation and assistive robots," *Acta Automatica Sinica*, vol. 44, no. 11, pp. 2000–2010, 2018.
- [2] D. Y. Y. Sim and C. K. Loo, "Extensive assessment and evaluation methodologies on assistive social robots for modelling human-robot interaction - a review," *Information Sciences*, vol. 301, pp. 305–344, 2015.
- [3] T. Min and S. Wang, "Research progress on robotics," *Acta Automatica Sinica*, vol. 39, no. 7, pp. 963–972, 2013.
- [4] M. Schumacher, J. Wojtusich, P. Beckerle, and O. von Stryk, "An introductory review of active compliant control," *Robotics and Autonomous Systems*, vol. 119, pp. 185–200, 2019.
- [5] O. C. Jr, Y. Hirata, and K. Kosuge, "Control of walking support system based on variable center of rotation," in *Proceedings of the IEEE International Conference on Intelligent Robots & Systems*, Sendai, Japan, October 2004.

- [6] P. K. Jamwal, S. Hussain, M. H. Ghayesh, and S. V. Rogozina, "Impedance control of an intrinsically compliant parallel ankle rehabilitation robot," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 6, pp. 3638–3647, 2016.
- [7] S.-Y. Jiang, C.-Y. Lin, K.-T. Huang, and K.-T. Song, "Shared control design of a walking-assistant robot," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 6, pp. 2143–2150, 2017.
- [8] W. Xu, J. Huang, Y. Wang, C. Tao, and L. Cheng, "Reinforcement learning-based shared control for walking-aid robot and its experimental verification," *Advanced Robotics*, vol. 29, no. 22, pp. 1463–1481, 2015.
- [9] W. X. Xu, J. Huang, Q. Y. Yan, Y. J. Wang, and C. J. Tao, "Research on walking-aid robot motion control with both compliance and safety," *Acta Automatica Sinica*, vol. 42, no. 12, pp. 1859–1873, 2016.
- [10] X. Han, M. Ge, J. Cui, H. Wang, and W. Zhuang, "Motion modeling of a non-holonomic wheeled mobile robot based on trajectory tracking control," *Transactions of the Canadian Society for Mechanical Engineering*, vol. 44, no. 2, pp. 228–233, 2019.
- [11] Q. Yan, W. X. Xu, J. Huang, and S. Cao, "Laser and force sensors based human motion intent estimation algorithm for walking-aid robot," in *Proceedings of the IEEE Annual International Conference on Cyber Technology in Automation Control and Intelligent Systems*, Shenyang, China, June 2015.
- [12] Y. Hirata, A. Muraki, and K. Kosuge, "Motion control of intelligent passive-type walker for fall-prevention function based on estimation of user state," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Orlando, FL, USA, May 2006.
- [13] K. T. Song and C. Y. Lin, "A new compliant motion control design of a walking-help robot based on motor current and speed measurement," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, St. Louis, MO, USA, October 2009.
- [14] K. T. Song and S. Y. Jiang, "Force-cooperative guidance design of an omni-directional walking assistive robot," in *Proceedings of the International Conference on Mechatronics & Automation*. IEEE, Beijing, China, August 2011.
- [15] J. Paulo, P. Peixoto, and U. J. Nunes, "ISR-AIWALKER: robotic walker for intuitive and safe mobility assistance and gait analysis," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 6, pp. 1110–1122, 2017.
- [16] C.-K. Lu, Y.-C. Huang, and C.-J. Lee, "Adaptive guidance system design for the assistive robotic walker," *Neurocomputing*, vol. 170, no. 25, pp. 152–160, 2015.
- [17] T. Degen, H. Jaeckel, M. Rufer, and S. Wyss, "SPEEDY: a fall detector in a wrist watch," in *Proceedings of the IEEE International Symposium on Wearable Computers*, White Plains, NY, USA, October 2003.
- [18] S. Qiu, Z. Wang, H. Zhao, and H. Hu, "Using distributed wearable sensors to measure and evaluate human lower limb motions," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 4, pp. 939–950, 2016.
- [19] C. P. Lee, A. W. C. Tan, and S. C. Tan, "Gait probability image: an information-theoretic model of gait representation," *Journal of Visual Communication and Image Representation*, vol. 25, no. 6, pp. 1489–1492, 2014.
- [20] W. Ye, Z. Li, C. Yang, J. Sun, C.-Y. Su, and R. Lu, "Vision-based human tracking control of a wheeled pendulum robot," *IEEE Transactions on Cybernetics*, vol. 46, no. 11, pp. 2423–2434, 2016.
- [21] M. Li, G. Xu, B. He, X. Ma, and J. Xie, "Pre-impact fall detection based on a modified zero moment point criterion using data from Kinect sensors," *IEEE Sensors Journal*, vol. 18, no. 13, pp. 5522–5531, 2018.
- [22] G. Lee, T. Ohnuma, N. Y. Chong, and S. Lee, "Walking intent-based movement control for JAIST active robotic walker," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 5, pp. 665–672, 2014.
- [23] X. Xi, H. Wu, J. Zuo, and Z. Luo, "Study on fall detection based on surface EMG and plantar pressure signal fusion," *Chinese Journal of Scientific Instrument*, vol. 36, no. 9, pp. 2044–2049, 2015.
- [24] J. Huang, W. X. Xu, S. Mohammed, and Z. Shu, "Posture estimation and human support using wearable sensors and walking-aid robot," *Robotics and Autonomous Systems*, vol. 26, pp. 24–43, 2014.
- [25] Y. Ma, R. Fallahzadeh, and H. Ghasemzadeh, "Glaucoma-specific gait pattern assessment using body-worn sensors," *IEEE Sensors Journal*, vol. 16, no. 16, pp. 6404–6415, 2016.
- [26] D. H. Zhao, J. Y. Yang, Y. N. Wang, and S. Y. Wang, "Recognition and analysis of abnormal gait in active walking rehabilitation training," *ICIC Express Letters, Part B: Applications*, vol. 8, no. 11, pp. 2185–2766, 2017.
- [27] D. Pei, Y. Hasegawa, and S. Nakagawa, "Fall detection and prevention control using walking-aid cane robot," *IEEE*, vol. 21, no. 2, pp. 625–637, 2016.
- [28] D. Pei, H. Jian, S. Nakagawa, K. Sekiyama, and T. Fukuda, "Fall detection and prevention in the elderly based on the ZMP stability control," in *Proceedings of the IEEE Workshop on Advanced Robotics and Its Social Impacts*, Tokyo, Japan, November 2013.
- [29] Q. Y. Yan, J. Huang, and Z. W. Luo, "Human-robot coordination stability for fall detection and prevention using cane robot," in *Proceedings of the 2016 International Symposium on Micro-Nano Mechatronics and Human Science*, Nagoya, Japan, November 2016.
- [30] K. Wakita, J. Huang, P. Di, K. Sekiyama, and T. Fukuda, "Human-walking-intention-based motion control of an omnidirectional-type cane robot," *IEEE*, vol. 18, no. 1, pp. 285–296, 2013.
- [31] M. Martin, E. Behan, and J. Chilleme, "Patient rehabilitation aid that varies treadmill belt speed to match a user's own step cycle based on leg length or step length," Biodex Medical Systems, Inc., US, 2003.
- [32] J. Yoon, H. S. Park, and D. L. Damiano, "A novel walking speed estimation scheme and its application to treadmill control for gait rehabilitation," *Journal of Neuroengineering and Rehabilitation*, vol. 9, no. 62, p. 62, 2012.
- [33] D. Zhao, J. Yang, M. O. Okoye, and S. Wang, "Walking assist robot: a novel non-contact abnormal gait recognition approach based on extended set membership filter," *IEEE Access*, vol. 7, pp. 76741–76753, 2019.
- [34] D. H. Zhao, J. Y. Yang, Y. N. Wang, and S. Y. Wang, "Fuzzy system based on the rule evolution strategy for directional intention identification of walking," *Chinese Journal of Scientific Instrument*, vol. 38, no. 11, pp. 2615–2625, 2017.
- [35] L. Wang, Y. Zhang, W. Peng, X. U. Bo, and Q. Wang, "SVR approach based on artificial bee colony optimization," *Systems Engineering and Electronics*, vol. 36, no. 2, pp. 326–330, 2014.
- [36] J. L. G. Nielsen, S. Holmgaard, N. Jiang, K. B. Englehart, D. Farina, and P. A. Parker, "Simultaneous and proportional force estimation for multifunction myoelectric prostheses using mirrored bilateral training," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 681–688, 2011.
- [37] A. Fujimoto, N. Matsumoto, Y. Jiang, S. Togo, and H. Yokoi, "Gait analysis-based speed control of walking assistive robot,"

- in *Proceedings of the IEEE International Conference on Intelligence and Safety for Robotics (ISR)*, Shenyang, China, August 2018.
- [38] C. H. Liu, C. Y. Chiang, P. Y. Lin, and Y. C. Chou, "A fall detection system using accelerometer and gyroscope," Master Thesis, Tatung University, 2011.
- [39] M. A. H. Farquad, V. Ravi, and S. B. Raju, "Churn prediction using comprehensible support vector machine: an analytical CRM application," *Applied Soft Computing*, vol. 19, pp. 31–40, 2014.
- [40] D. H. Zhao, J. Y. Yang, D. C. Bai, and Y. L. Jiang, "Transfer method of multiple welfare-robots based on minimal fuzzy system," *Robot*, vol. 41, no. 6, pp. 813–822, 2019.
- [41] S. Ding, X. Ouyang, T. Liu, Z. Li, and H. Yang, "Gait event detection of a lower extremity exoskeleton robot by an intelligent IMU," *IEEE Sensors Journal*, vol. 18, no. 23, pp. 9728–9735, 2018.
- [42] B. Zhong, W. Niu, E. Broadbent, A. Mcdaid, T. M. C. Lee, and M. Zhang, "Bringing psychological strategies to robot-assisted physiotherapy for enhanced treatment efficacy," *Frontiers in Neuroscience*, vol. 13, p. 984, 2019.
- [43] F. Dehais, E. A. Sisbot, R. Alami, and M. Causse, "Physiological and subjective evaluation of a human-robot object hand-over task," *Applied Ergonomics*, vol. 42, no. 6, pp. 785–791, 2011.
- [44] Y. Hu, M. Benallegue, G. Venture, and E. Yoshida, "Interact with me: an eExploratory study on interaction factors for active physical human-robot interaction," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6764–6771, 2020.
- [45] S. H. Fairclough, "Fundamentals of physiological computing," *Interacting with Computers*, vol. 21, no. 1-2, pp. 133–145, 2009.
- [46] M. J. Khan and K.-S. Hong, "Passive BCI based on drowsiness detection: an fNIRS study," *Biomedical Optics Express*, vol. 6, no. 10, pp. 4063–4078, 2015.

Research Article

Dynamic Invariant-Specific Representation Fusion Network for Multimodal Sentiment Analysis

Jing He , Haonan Yanga , Changfan Zhang , Hongrun Chen , and Yifu Xua

College of Electrical and Information Engineering, Hunan University of Technology, Zhuzhou 412007, China

Correspondence should be addressed to Changfan Zhang; zhangchangfan@263.net

Received 29 November 2021; Revised 31 December 2021; Accepted 6 January 2022; Published 24 January 2022

Academic Editor: Zhongxu Hu

Copyright © 2022 Jing He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multimodal sentiment analysis (MSA) aims to infer emotions from linguistic, auditory, and visual sequences. Multimodal information representation method and fusion technology are keys to MSA. However, the problem of difficulty in fully obtaining heterogeneous data interactions in MSA usually exists. To solve these problems, a new framework, namely, dynamic invariant-specific representation fusion network (DISRFN), is put forward in this study. Firstly, in order to effectively utilize redundant information, the joint domain separation representations of all modes are obtained through the improved joint domain separation network. Then, the hierarchical graph fusion net (HGFN) is used for dynamically fusing each representation to obtain the interaction of multimodal data for guidance in the sentiment analysis. Moreover, comparative experiments are performed on popular MSA data sets MOSI and MOSEI, and the research on fusion strategy, loss function ablation, and similarity loss function analysis experiments is designed. The experimental results verify the effectiveness of the DISRFN framework and loss function.

1. Introduction

Multimodal sentiment analysis (MSA), as an emerging field of natural language processing (NLP), aims to infer the speaker's emotion by exploring clues in multimodal information [1–3]. Many methods in MSA focus on exploring the complex fusion mechanism to improve the performance of MSA [4–6]. However, these fusion technologies present a bottleneck due to the difficulty in obtaining interaction between heterogeneous modes. The common method to solve this problem is to map the heterogeneous feature to the common subspace in the representation learning process [7]. However, some unique features of each mode are ignored by those methods. These unique features can be used as complementary information between modes. Effective use of this complementary information can help the network improve performance. For this consideration, this paper intends to use supplementary information on the basis of shared representation. And then, a dynamic fusion mechanism is established to fuse the modal features to

obtain the interactive information. This study mainly aims to explore a sentiment analysis framework based on multimodal representation learning and the dynamical fusion method.

For multimodal representation learning methods, since multimodal data is usually a sequence with different feature dimensions, long-short memory neural network (LSTM) is a powerful tool to deal with such problems [8]. Therefore, different LSTMs are used to extract features of different modalities in many methods, such as memory fusion network (MFN) [9], graph-memory fusion network (Graph-MFN) [10]. However, a single LSTM is difficult to apply to the feature distribution of each mode at the same time. Therefore, there are studies using different networks to represent different modal information, such as tensor fusion network (TFN) [11], low-rank multimodal fusion net (LMF) [12]. It is worth mentioning that the information between modalities was not used fully before fusion in these methods. The shared features and special features of two data sources are captured by domain separation network (DSN) using adversarial learning and soft orthogonal constraint [13]. And then, these features are used to perform domain adaptive

tasks. The combination of shared features and special features can effectively solve the problem that the redundant information between different data sources is not fully utilized. In other words, the DSN is improved and adopted to perform multimodal sentiment analysis tasks in this paper. It is named improved joint domain separation network (improved JDSN).

In this paper, the improved JDSN is adopted to learn the joint representation of modality-invariant and modality-specific of all modes in the common-special subspace. The former aims to map all the modes of discourse to the common subspace to shorten the distance between modes to effectively reduce the extra burden of fusion work. The latter aims to extract special representation from each mode as complementary information. Then, the combination of two representations can fully use the complementary information between modes. In addition, the modal interactions were mostly obtained by feature connection fusion in early work [14]. However, these methods are unable to dynamically adjust the contribution of each mode in the fusion process. Mai et al. assumed that the multimodal fusion process is a hierarchical interactive learning process [15, 16] and designed a ARGF network to solve the problem [15]. The ARGF was comprised of two stages: a joint embedding space learning stage and a hierarchical graph fusion net (HGFN) stage. In the HGFN stage, firstly, the unimodal dynamic layer, bimodal dynamic layer, and trimodal dynamic layer are modelled, and then the outputs of each dynamic layer are connected to obtain the interaction features of each mode. However, the method of joint embedding space learning also has a problem that the redundant information was not fully utilized. Therefore, the improved JDSN and HGFN are combined to optimize the network’s ability to capture modal interactions by rationally using redundant information in this paper.

In summary, firstly, the applied DSN in this paper is improved in the aspects of the following: (1) The mode of DSN is extended; (2) The orthogonal constraint loss between special representations of different modes is additionally considered (See Section 3.3.1); (3) Adversarial loss is replaced by a more advanced similarity metric (CMD) (See Section 3.3.2); (4) Invariant and specific representation are jointed at the output of the network (see Section 3.2.3). Then, combining the improved JDSN and HGFN, a new framework (DISRFN) is proposed in this paper to deal with MSA problems. The main contributions are as follows:

- (1) A multimodal sentiment analysis framework (DISRFN) is proposed in this study. It can perform the fusion of various representations dynamically while emphasizing learning invariant and specific joint representations of various modes.
- (2) A new loss function is designed, which can improve the effect of semantic fusion clustering whilst assisting the model in learning the target subspace representation effectively.
- (3) The performance analysis experiments of MSA tasks is designed on the benchmark data sets MOSI and

MOSEI. The results confirm the advancement of the DISRFN model and fusion strategy, the effectiveness of the loss function, and the rationality of similarity loss function selection.

The remainder of this paper consists of the following parts. In Section 2, the correlation work is briefly reviewed. Section 3 introduces the structure of the DISRFN model and the proposed learning method in detail. Section 4 explains the experimental details, parameter settings, and network component design. The experimental results are analyzed in Section 5. Section 6 shows the summary and prospects.

2. Correlation Work

In multimodal sentiment analysis, the mainstream multimodal learning methods include multimodal fusion representation and multimodal representation learning, which will be discussed in this section.

2.1. Multimodal Fusion Representation. In recent years, some complex and efficient fusion representation mechanisms have been gradually proposed. Amir Zadeh et al. put forward TFN to obtain the trimodal fusion representations by using the outer product [11]. On this basis, a low-rank multimodal fusion net (LMF) was proposed. This network performs multimodal data fusion employing a low-rank tensor and obtains better results [12]. Mai et al. proposed a strategy “divide and rule, unite many into one” to transfer local tensor and global fusion, which was extended in multiconnected bidirectional long-short time memory network (Bi-LSTM) [17, 18]. In addition to the tensor fusion method, the recursive fusion method has been developed better. For example, a recursive multilevel fusion network (RMFN) is used for specialized and effective fusion through decomposing the fusion problems into several parts [19]. The more attention-based recursive network (MARN) is used to fuse cyclic memory representations of different modes of long-short term hybrid memory networks (LSTHM) by using a more attention block [20]. Hierarchical polynomial fusion network (HPFN) is used to recursively integrate and transfer the local correlation to the global correlation through multilinear fusion [21]. Moreover, the multiview learning method plays an important role in multimodal fusion [22]. For example, MFN designed by Amir Zadeh et al. is used to fuse the memory of different modes of LSTM system based on incremental attention memory network (DAMN) and gated memory network (MVGN) [9], and it is successfully used to solve multiview problems. Furthermore, to analyze the explainability of MFN, the dynamic fusion graph model (DFG) is embedded into MFN, and a Graph-MFN obtained finally has excellent performance and is explainable [10]. Recently, word-level fusion representation has also been a wide concern [23]. For example, a repeated participation variation network (RAVEN) is used to model multimodal language through work representation transfer based on facial expression [24]. Chen et al. modeled the time-dependent multimodal dynamics through cross-modal work alignment [25]. However, most

of these methods use complex fusion mechanisms or add additional fusion modules, which will increase the amount of calculation and slow down the speed of network convergence. In contrast, this paper uses a hierarchical mechanism to model the dynamics of each fusion layer, which can quickly fuse the information of each mode.

2.2. Multimodal Representation Learning. Multimodal representation learning is mainly divided into two types, namely, common subspace representations and factorised representations. The two types of study on common subspace representations amongst modes are the correlation-based model and adversarial learning-based model. In terms of a correlation-based model, Shu et al. proposed an extensible multilabel canonical correlation analysis (sml-CCA) for cross-modal retrieval [26]. Kaloga et al. proposed a multiview graph canonical correlation analysis based on variational graph neural network for classification and clustering tasks [27]. Verma et al. proposed a deep network with high-order information and single sequence information (Deep-HOSeq) for fusing multimodal sentiment data [28]. Mai et al. learned the embedding space within invariant mode based on a new encoding-decoding classifier framework in confrontation [15]. Pham et al. proposed a robust joint representation method to learn by shifting between modes under the constraints of cyclic consistency loss [29]. In terms of the adversarial learning-based model, Wu and Qiang et al. proposed the generative adversarial net based on specific mode and sharing and the adversarial hashing algorithm based on deep semantic similarity, respectively, to obtain cross-modal invariance [30, 31]. However, these methods only learn about the shared representation of the model and lack the consideration of the special representation of the modal. For factorized representations, Amir Zadeh et al. proposed a multimodal factorized model (MFM) to factorize multimodal representations into multimodal discriminant factor and multimodal special generation factor [32]. Liang et al. proposed a multimodal baseline model (MMB) to learn the cases of multimodal embedding based on the factorized method [33]. Wang et al. proposed a joint and separate matrix factorized hashing method, which could be used to learn common and specific attributes of multimodal data at the same time [34]. Fang et al. proposed a new semantic enhanced discrete matrix factorized hashing (SDMFN), which could directly extract the common hashing representation from the reconstructed semantic polynomial similar graph, causing the hash code to be more discriminative [35]. Caicedo et al. proposed a multimodal image representation based on nonnegative matrix factorisation to synthesise visual features and text features [36]. However, most of these factorized methods adopt the form of matrix decomposition, which may have the problem of incomplete feature representation. In contrast, the improved JDSN designed in this paper can obtain a richer shared-special representation of each mode in a simpler way.

3. The Proposed Method

3.1. Task Setting. In general, the proposed framework is mainly used to study the trimodal data. Figure 1 shows the flowchart of the proposed multimodal fusion framework. This framework consists of two parts, as follows: (1) improved JDSN for learning trimodal data-specific shared subspace joint representation; (2) HGFN for fusing trimodal joint representation, thereby realizing dynamical effective semantic clustering. This study introduces this network framework in the following section.

Moreover, the discourse data are divided into N sequences composed of segment S to facilitate detecting emotion in video by using multimodal data. Each segment S includes three low-level feature sequences in linguistic (l), visual (v), and auditory (a) modes. These feature sequences are represented as $S_l \in \mathbb{R}^{t_l \times d_l}$, $S_v \in \mathbb{R}^{t_v \times d_v}$, $S_a \in \mathbb{R}^{t_a \times d_a}$. Amongst them, t_m and d_m ($m \in \{l, v, a\}$) represent the length of discourse and the dimension of the corresponding feature, respectively. Given this data sequence, the study aims to predict the emotional state of the predefined set. This emotional state is a continuous dense variable $y \in \mathbb{R}$. In addition, to effectively use multimodal data, linguistic (l), visual (v), and auditory (a) trimodal feature sequences, they should be aligned with emotional state label y .

The framework of DISRFN is shown in Figure 1: (1) The data of the three modes are fed into the corresponding Bi-LSTM and BERT models to obtain the discourse-level feature representations; (2) The discourse-level feature representations of each mode are fed into the corresponding MLP to obtain the representation of unified dimension; (3) The unified representations of each mode are fed into the corresponding encoder and shared encoder to obtain the shared representations and special representations; (4) The shared representations are added with a special representation of each modal to obtain the joint domain separation representations; (5) The joint domain separation representations of each mode are fed into the corresponding decoder to obtain the reconstruction loss; (6) The joint domain separation representations of each mode are fed into HGFN for dynamic fusion to perform MSA task.

3.2. Dynamic Invariant-Specific Representation Fusion Network

3.2.1. Discourse-Level Feature Representation. Firstly, the stacking bidirectional long-short time memory neural network (sLSTM) is used to map the feature sequence (S_v , S_a) in visual (v) and auditory (a) modes to obtain the underlying features of the sequence. Its output includes the hidden representations of LSTM end state, namely, F_v and F_a , as follows:

$$\begin{aligned} F_v &= \text{sLSTM}(S_v; \theta_v^{\text{LSTM}}), \\ F_a &= \text{sLSTM}(S_a; \theta_a^{\text{LSTM}}), \end{aligned} \quad (1)$$

where θ_v^{LSTM} and θ_a^{LSTM} refer to the parameters of sLSTM on visual and auditory modes.

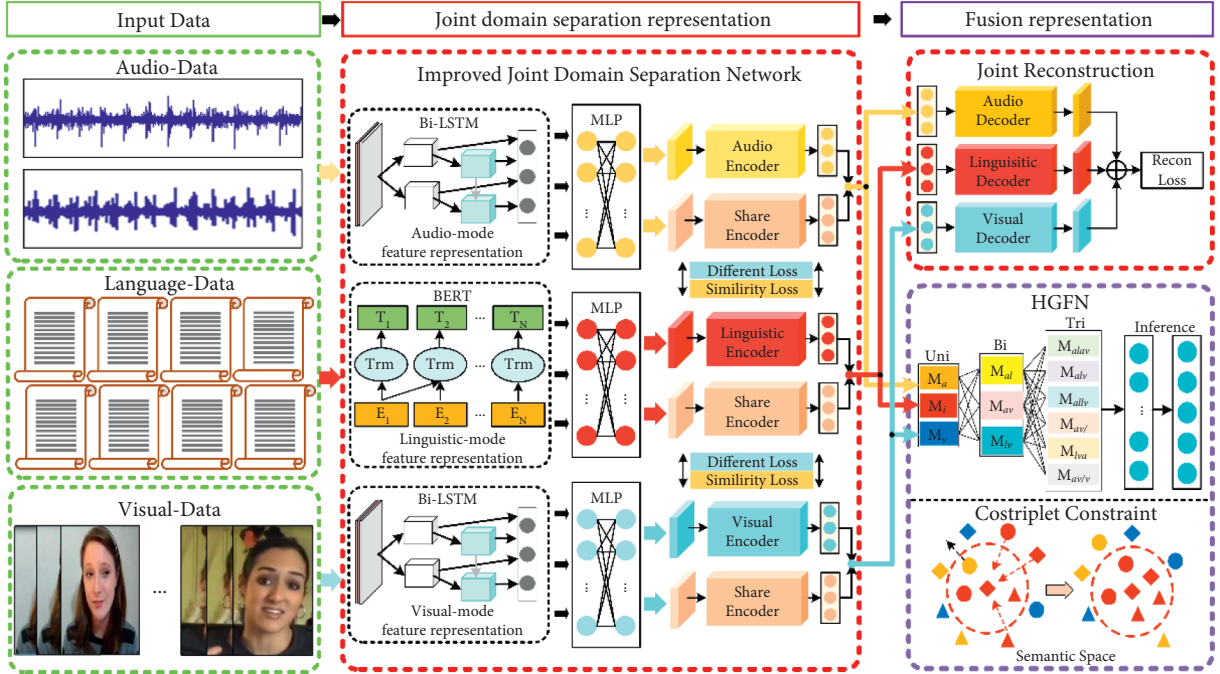


FIGURE 1: The framework of DISRFN. Note: Bi-LSTM: bidirectional short and long memory network; BERT: bidirectional encoder representation from transformers; MLP: multilayer perceptron; audio encoder (decoder): encoder (decoder) of auditory mode; linguistic encoder (decoder): encoder (decoder) of linguistic mode; visual encoder (decoder): encoder (decoder) of visual mode; share encoder: shared encoder of three modes; HGFN: hierarchical graph fusion net.

Secondly, for the text feature sequence (S_l) in linguistic mode, most linguistic features are embedded through Glove [37]. However, in recent studies [38], such as the advanced ICCN [39] model, the pretraining BERT model is used as the feature extractor of text discourse. A better result than the Glove method is obtained. Therefore, the feature representation F_l of text is obtained through the pretraining BERT model, as follows:

$$F_l = \text{BERT}(S_l; \theta_l^{\text{BERT}}), \quad (2)$$

where θ_l^{BERT} refers to the parameter of the BERT model.

3.2.2. Unified Representation of Features. The dimensions of discourse-level features are different. In order to facilitate the encoding-decoding operation in the back-end network, multilayer perceptron (MLP) is used to unify mapping these features to O_m , as follows:

$$O_m = \text{MLP}(F_m; \theta_m^{\text{MLP}}), \quad (m \in \{l, v, a\}), \quad (3)$$

where θ_m^{MLP} refers to a parameter of multilayer perceptron networks in different modes; MLP consists of dense connection layers and a normalized layer activated by relu function.

3.2.3. Improved Joint Domain Separation Representation. In this part, based on the improved JDSN, the unified mapping representation of each mode is factorized into two parts, namely, modality-invariance and modality-specificity. Amongst them, the sharing encoder E^c is used to learn

invariant representation in the common subspace to narrow the gap in the heterogeneity between modes [40]. The specific encoder E_m^p is used to capture the specific representation in a specific subspace. The process is as follows.

Firstly, after obtaining the unified mapping vector O_m of each mode, the mode-sharing encoder E^c (weight sharing) is used to obtain modality-invariant representation (h_m^c), and the mode-specific encoder E_m^p is used to extract modality-specific representation (h_m^p), as follows:

$$h_m^c = E^c(O_m; \theta^c), h_m^p = E_m^p(O_m; \theta_m^p), \quad (m \in \{l, v, a\}), \quad (4)$$

where θ^c refers to a parameter of mode-sharing encoder; θ_m^p refers to a parameter of mode-specific encoder; E^c has the same structure as that of E_m^p , which is composed of a dense connection layer activated by sigmoid function.

Then, hidden layer vectors h_m^p and h_m^c are generated through feedforward propagation of neural network, and the joint domain separation representation is obtained through vector addition "+", as follows:

$$h_m = h_m^c + h_m^p, \quad (m \in \{l, v, a\}), \quad (5)$$

where h_m refers to the joint domain separation representation of mode m , and it has the feature representation of shared subspace and specific subspace characteristics.

3.2.4. Hierarchical Graph Fusion Representation. After obtaining the joint domain separation representation of each mode, it is necessary to fuse each representation to obtain the interaction information of each mode.

As shown in Figure 2, HGFN is composed of three dynamic layers (unimodal dynamic layer, bimodal dynamic layer, and trimodal dynamic layer). Unimodal dynamic layer is modeled by self-attention weighting each unimodal information vector. Bimodal dynamic layer is modeled by weighting bimodal information vectors (e.g., M_{al}) using the correlation weight between unimodal vectors. Trimodal dynamic layer is constructed through weighting trimodal information vectors (e.g., M_{alv} or M_{allv}) by the correlation weight between unimodal vectors. Finally, three dynamic layers are used for vector connection and fusion to realize the dynamic fusion of multimodal features in HGFN. This hierarchical modeling method is more conducive to exploring the interaction between modes [12]. Therefore, HGFN, which can preserve all modal interactions, is introduced to fuse the obtained joint domain separation representations of different modes to explore multimodal interaction in this section. The fusion representation is as follows:

$$\text{Fusion} = \text{HGFN}(h_l, h_v, h_a; \theta^{\text{HGFN}}), \quad (6)$$

where ‘‘Fusion’’ refers to the output of HGFN; θ^{HGFN} refers to the parameters of HGFN. Then, the predictive neural network (P) is used for prediction, as follows:

$$\text{Pred} = P(\text{Fusion}; \theta^{\text{Pre}}), \quad (7)$$

where ‘‘Pred’’ refers to the output of the predictive network; ‘‘P’’ refers to a predictive network, including a standardized layer and the fully connected layers; θ^{Pre} refers to the parameter of the predictive network. Moreover, the specific parameters of the model are described in the experimental section.

3.3. Learning Process. A joint loss function is newly set to effectively learn the network model, as follows:

$$L_{\text{total}} = L_{\text{task}} + \alpha L_{\text{diff}} + \beta L_{\text{sim}} + \gamma L_{\text{recon}} + \eta L_{\text{trip}}, \quad (8)$$

where α , β , γ , and η refer to weights of the interaction. They determine the contributions of each loss L_{diff} , L_{sim} , L_{recon} , and L_{trip} to total loss L_{total} . In addition, each loss is analyzed and introduced in the remaining section.

3.3.1. Differential Loss. Some studies have shown that a nonredundant effect can be achieved by applying soft orthogonality constraint to two representation vectors [13, 41]. Therefore, the constraint is used to drive the sharing-encoder E^c and specific-encoder E_m^p to perform encoding representation to different aspects, that is, modality-invariant and modality-specific representations. Soft orthogonality constraint is defined as follows.

When training a batch of data, H_m^c and H_m^p are set as the two matrices, respectively. The rows of the two matrices correspond to invariant representation h_m^c and specific representation h_m^p of mode m in each batch of data, respectively. The orthogonality constraint of the modal vector is calculated as follows [13]:

$$L_{\text{diff}} = \sum_{m \in \{l, v, a\}} \|H_m^{cT} H_m^p\|_F^2 + \sum_{\substack{(m_1, m_2) \in \{(l, a), \\ (l, v), (a, v)\}}} \|H_{m_1}^{pT} H_{m_2}^p\|_F^2, \quad (9)$$

where $\|\cdot\|_F^2$ refers to squared Frobenius norm.

3.3.2. Similarity Loss. Similarity loss (L_{sim}) used to constrain shared subspace can reduce the difference in the heterogeneity between the shared representations of different modes [42]. Central moment discrepancy (CMD) is used to measure the difference between two distributions by matching order-wise moment differences of two representations [43]. Compared with other methods (e.g., MMD and DANN), it is a more efficient and concise distance measurement. Therefore, CMD is selected as the similarity loss in this paper. It is defined as follows.

X and Y are set as bounded random samples with probability distributions p and q in a compact interval $[a, b]^N$, respectively. CMD is defined as follows [43]:

$$\text{CMD}(X, Y) = \frac{1}{|b-a|} \|E(X) - E(Y)\|_2 + \sum_{k=2}^K \frac{1}{|b-a|^k} \|C_k(X) - C_k(Y)\|_2 \quad (10)$$

$$C_k(X) = E((x - E(X))^k)$$

$$E(X) = \frac{1}{|X|} \sum_{x \in X} x,$$

where $E(X)$ refers to the empirical expectation vector of sample X ; $C_k(X)$ refers to the vector of all k -order sample centre moments in the X coordinate.

In this paper, the similarity loss is calculated by summing the CMD distances of the shared representations of every two modes. Its representation is as follows:

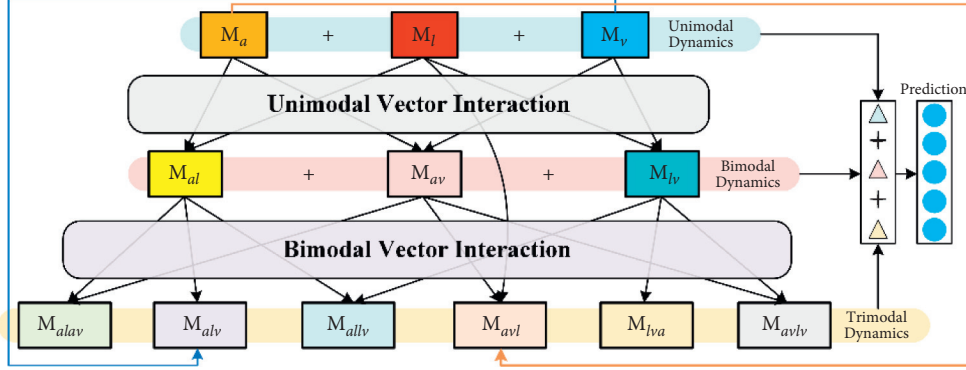


FIGURE 2: The framework of HGFN.

$$L_{\text{sim}} = \sum_{\substack{(m_1, m_2) \in \{(l, a), \\ (l, v), (a, v)\}}} \text{CMD}(h_{m_1}^c, h_{m_2}^c), \quad (11)$$

Moreover, the reason for selecting CMD as the similarity loss will be discussed in Experimental part 5.4.

3.3.3. Reconstruction Loss. When soft orthogonality constraint is enforced, the risk of specific encoder learning trivial representation exists. However, the reconstruction loss can be added to ensure that the encoder can capture the details of each mode to solve these problems [13]. Initially, the modal decoder D_m is used to reconstruct the joint domain separation representation vector h_m of mode m , and the output of reconstruction is \hat{h}_m . Then, the reconstruction loss is represented by the mean square error loss between h_m and \hat{h}_m , as follows [13]:

$$L_{\text{recon}} = \frac{1}{3} \left(\sum_{m \in \{l, v, a\}} \|h_m - \hat{h}_m\|_2^2 \right), \quad (12)$$

where $\|\cdot\|_2^2$ refers to squared L_2 -norm.

3.3.4. Cosine Triplet-Margin Loss. In the fusion representation of joint domain separation representation vector, to ensure the high-level relationship of the similarity between all projects, the representation distance of discourse segments with similar semantics between different modes is minimized through cosine triplet-margin loss L_{trip} , and the distance between different discourse segments is maximized [44].

For example, in linguistic and visual modes, a triple representation (h_l, h_v^+, h_v^-) is established. Amongst them, visual representation h_v^+ is positively correlated with linguistic representation h_l in semantics. At the same time, visual representation h_v^- is the contrary. Therefore, the cosine triplet-margin loss of linguistic mode is shown as follows [44]:

$$L_{\text{trip}}^l = \sum_{m \in \{v, a\}} \max(\cos(h_l, h_m^-) - \cos(h_l, h_m^+) + \text{margin}, 0), \quad (13)$$

where h_m^+, h_m^- refers to the joint domain separation representation vector of mode m ; “margin = 1” is a boundary parameter.

In the same way, the cosine triplet-margin loss of visual mode and auditory mode can be described as follows:

$$L_{\text{trip}}^v = \sum_{m \in \{l, a\}} \max(\cos(h_v, h_m^-) - \cos(h_v, h_m^+) + \text{margin}, 0), \quad (14)$$

$$L_{\text{trip}}^a = \sum_{m \in \{l, v\}} \max(\cos(h_a, h_m^-) - \cos(h_a, h_m^+) + \text{margin}, 0). \quad (15)$$

Based on formulas (13)–(15), the total cosine triple margin loss is represented as follows:

$$L_{\text{trip}} = L_{\text{trip}}^l + L_{\text{trip}}^v + L_{\text{trip}}^a. \quad (16)$$

3.3.5. Task Loss. The mean square error (MSE) is used as the task loss of the network to predict continuous dense variables. For N_b discourse data in one batch, this loss calculation is as follows:

$$L_{\text{task}} = \frac{1}{N_b} \sum_{i=0}^{N_b} \|y_i - \hat{y}_i\|_2^2. \quad (17)$$

where y_i refers to the actual emotional label; \hat{y}_i refers to the predictive value of the network.

4. Experiment

In this section, the required data sets, evaluation index, and experimental details (experimental environment, experimental parameters, and network structure) are described.

4.1. Datasets. The data set is introduced in this section. This data set includes two parts, namely, CMU-MOSI and CMU-MOSEI.

CMU-MOSI data set: this data set is a collection of monologues on YouTube, including videos with 93 comments from different speakers. These common videos consist of 2199 subjective discourses. These discourses are manually

marked with continuous opinion scores in the range of -3 to 3 . Amongst them, $-3/+3$ represents strong negative/positive emotions. A total of 1283 segment samples are used for training, 229 segments are used for verification, and 686 segments are used for testing.

CMU-MOSEI data set: it is an improved version of MOSI; it includes 23453 annotated discourse segments, which are from 5000 videos, 1000 different speakers, and 250 different topics. A total of 1283 segment samples are still used for training, 229 segments are used for verification, and 686 segments are used for testing.

The problems on multimodal signal (linguistic, visual, and auditory) acquisition and modal data pretreatment are solved based on CMU-Multimodal SDK¹ in many studies [45]. This tool library is a machine learning platform used for developing high-level multimodal models and acquiring and processing multimodal data by Amir Zadeh et al. It integrates the acquisition and alignment method of benchmark data sets (MOSI and MOSEI). Similarly, this tool library is used to solve the problems of data acquisition and alignment.

4.2. Evaluation Index. This experiment is a regressive task. Therefore, the mean absolute error (MAE) and Pearson correlation coefficient (Corr) are adopted to measure the test results. In addition, the classification index is considered in the experiment, including five-classification accuracy (Acc-5) in affection domain ($-2,2$), two-classification accuracy (Acc-2) including positive and negative emotion (p/g), and F1score (F1-Score).

4.3. Experimental Settings. This method is tested on Pytorch in this section. The grid searching of hyperparameter is performed in a data verification set to identify appropriate hyperparameter, and the best model and hyperparameter are saved. In grid searching, limited option sets for setting hyperparameters are as follows: $\alpha \in \{0.3, 0.4\}$, $\beta \in \{0.7, 0.8, 0.9, 1.0\}$, $\gamma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$, $\eta \in \{0.01, 0.1\}$ and $\text{drop} \in \{0, 0.1, 0.2, 0.3, 0.4\}$; the hidden layer sizes of the representation and predictive network can be reviewed from the following: $\text{Hid} \in \{128, 256\}$, $\text{P_h} \in \{50, 64\}$.

In the iterative optimization process, Adam optimizer with max epoch = 20, batch_size = 16, and learning rate of 0.0001 are used to train the network. The grid searching results of all data sets are shown in Table 1, and based on the hyperparameter settings, Figure 3 shows the model component structural diagram. Note: (1) FC Layer is the dimension of the fully connected layer; (2) LSTM is the dimension of the LSTM hidden layer; (3) Layer-Norm is a dimension of the batch normalization layer; (4) Dropout is the rate of dropout; (5) BERT is the output dimension of the BERT model; (6) Hid/drop/P_h is hyperparameters.

4.4. Experimental Process. This section mainly introduces the experimental process, the specific experimental steps are as follows:

TABLE 1: Hyperparameter settings in this article.

Hyperparameter	MOSI	MOSEI
CMD K	5	5
Batch_size	16	16
α	0.3	0.4
β	1.0	0.8
γ	0.4	0.4
η	0.1	0.01
Drop	0.4	0.1
Hid	256	256
P_h	64	50

- (1) Manual feature extraction of video and audio: for CMU-MOSI and CMU-MOSEI, Facet² and COVAREP [46] are used to extract the manual features of visual and auditory sequences. Amongst them, the dimensions d_v of the visual feature are 47 and 35, respectively, and the dimension d_a of the auditory feature is 74.
- (2) Discourse-level feature extraction: for linguistic mode, because the BERT model has text embedding and representation functions, the pretraining model of BERT is directly used to extract linguistic features. Its discourse-level feature is represented as feature representation F_l with dimension of 768 [47]. And then, visual and auditory features at the discourse-level F_v and F_a are obtained based on sLSTM.
- (3) Unified representation mapping: MLP is adopted to map linguistic, visual, and auditory representation vectors F_l , F_v , and F_a to an output O_m with the unified dimension size.
- (4) Improved joint domain separation representation: O_m is input to sharing encoder and specific encoder to obtain hidden layer representation h_m^c, h_m^p . And then, an improved joint domain separation representation h_m is obtained through vector addition ($h_m^p + h_m^c$).
- (5) Fusion inference: the joint domain separation representation vector is sent to the HGfN to perform fusion and prediction tasks.
- (6) Calculating loss function and training: loss function is calculated to train the neural network and make cyclic iteration.

5. Results and Analysis

Model comparison experiments, research on fusion strategy, research on loss function ablation, and research on similarity loss selection are designed in this section. All experiments are discussed by combining visualization and quantitative analysis.

5.1. Model Comparison Experiments Result. In the comparison experiment, some classical models (TFN, LMF, MFN, Gragh-MFN, MARM, and MISA) are reproduced. In addition, some derived fusion model based on LSTHM [17]

Private _ Encoder – E_m^p		Share _ Encoder – E^c		Decoder – D_m	
Private Encoder	FC Layer:Hid	Share Encoder	FC Layer:Hid	Decoder	FC Layer:Hid
	Sigmoid ()		Sigmoid ()		Sigmoid ()
Visual _ sLSTM		Acoustic _ sLSTM		Language – BERT	
sLSTM MLP	LSTM:47	sLSTM MLP	LSTM:74	BERT MLP	BERT:768
	Layer-Norm:47		Layer-Norm:74		FC Layer:Hid
	LSTM:47		LSTM:74		Relu ()
	FC Layer:Hid		FC Layer:Hid		Layer-Norm:Hid
	Layer-Norm:Hid		Layer-Norm:Hid		
Attention – MAN		Graph _ Fusion – MLF		Prediction – P	
Attention Block	FC Layer:Hid	Graph fusion Block	FC Layer:2*Hid	Prediction Networks	Layer_Norm:3*Hid
	FC Layer 1		Leaky Relu ()		Dropout:drop
	Sigmoid ()		FC Layer:64		FC Layer:3*Hid
	FC Layer:Hid		Tanh ()		
	Tanh ()		FC Layer:P_h		
		Tanh ()			
		FC Layer:1			

FIGURE 3: The parameter setting of modules.

is designed to comparison with the proposed framework (DISRFN). The result is shown in Tables 2 and 3.

Tables 2 and 3 show that our method achieves the best performance under two data sets. That is, it exceeds the comparison model in terms of MAE, Corr, Acc, and other comprehensive indexes. These results show that the proposed model exceeds some complex fusion mechanisms (e.g., TFN, MFN, and Gragh-MFN) in the performance. The reason is that these methods ignore the exploration of modal invariant space while the proposed method obtains a joint representation of invariant-specific space.

Moreover, it can be seen from the ‘‘CPU Clock’’ items in Tables 2 and 3. Compared with the model that also applies the mechanism fusion (TFN, LMF, MFN, Gragh-MFN, MARM, ARGF, LSTHM-DFG, LSTHM-Out Product), the proposed method is at a disadvantage in the aspect of real-time due to the relatively large number of parameters in the representation learning. However, compared with the model that uses additional networks in the fusion part (MISA, LSTHM-AttFusion, LSTHM-Concat), the proposed method has an advantage when it comes to real-time. Therefore, compared with the baseline model, the proposed method has moderate real-time performance when the various MSA indicators are optimal.

In Section 3.2.1, the reason for using the BERT pretraining model to extract discourse-level features of language modality instead of Glove method is explored. Tables 2 and 3 show that, compared with the baseline model based on the Glove word embedding method, and LSTHM-derived fusion model, various evaluation indexes are improved significantly by the model using BERT (DISRFN and MISA). It proves that the application of the BERT method is reasonable. Moreover, compared with the MISA model using BERT, the proposed model still has a slight advantage. The difference is probably caused by different fusion strategies. The comparative experiment is carried out in the next section to further discuss the effectiveness of the fusion strategy of this model.

5.2. Fusion Strategy Comparison Result. In this section, a fusion strategy comparison experiment is designed in the MOSI data set to verify the effectiveness of the HGFN fusion

strategy. The improved JDSN component remains unchanged in the experiment, and the fusion component is replaced with Multi-Attention Fusion (AttFusion), vector concatenation fusion (Concate), dynamic fusion net (DFN), and other strategies. Then, the results are concluded, as shown in Table 4.

The results shown in Tabel 4 indicate that HGFN has a significantly improved performance compared with other fusion methods. The reason for these results is that HGFN not only models single-modal, bimodal, and trimodal layers dynamically but also obtains trimodal fusion representations more comprehensively by the splicing mode of various modal layers. Moreover, to verify the dynamicity of the graph fusion network, the weight change of the fusion process is visualized as follows.

As shown in Figure 4, the vertical axis represents the iteration order, and the horizontal axis represents the interaction information vector in the dynamic layer. The value in the figure represents the weight of the corresponding information vector. The results of vertical axis analysis indicate that the contributions of different discourse segments to the same modal interaction information vector are almost unchanged. The reason is that the modal data are affected by the similarity constraint in the domain separation representation learning prior to fusion, which reduces the fluctuation in the difference amongst all sample representations. Through the observation of the horizontal axis, for single-modal vector weight (the first three columns), the contributions of linguistic mode to the prediction result are the most evident. The reason is that language text is usually the most important information in MSA. For bimodal vector weight (fourth–sixth column), weight ‘‘tv’’ is closer to ‘‘ta’’ and significantly greater than weight ‘‘va’’. The reason may be that linguistic mode plays a more important role in bimodal fusion than other modes. Through observation of the trimodal vector weight (the seventh–twelfth column), the vector weight obtained by fusing one bimodal vector and one single-modal vector is close to 0. However, the vector weight obtained by fusing two bimodal vectors is dominant in the trimodal information. It indicates that modeling the interaction process of every two bimodal vectors is

TABLE 2: Comparison experiments of multimodal models in MOSI

Model	MAE	Mul_Acc2	Mul_Acc5	Corr	F1_Score	CPU_Clock
TFN [11]	1.016	0.765	0.386	0.604	0.765	0.404
LMF [12]	1.009	0.767	0.362	0.604	0.769	0.395
MFN [9]	1.007	0.773	0.329	0.632	0.773	0.379
ARGF [15]	0.857	0.814	0.423	0.712	0.815	0.147
Gragh-MFN [10]	1.003	0.784	0.360	0.623	0.785	0.454
MARM [20]	1.028	0.756	0.351	0.625	0.755	0.345
LSTHM [20]-AttFusion	1.087	0.745	0.375	0.608	0.744	1.527
LSTHM [20]-Concat	1.056	0.750	0.370	0.581	0.752	1.524
LSTHM [20]-DFG	0.992	0.758	0.401	0.626	0.757	0.357
LSTHM [20]-Out_Product	1.092	0.764	0.332	0.569	0.764	0.708
MISA [41]	0.827	0.819	0.440	0.726	0.819	0.839
DISRFN (ours)	0.798	0.834	0.468	0.734	0.836	0.737

TABLE 3: Comparison experiments of multimodal models in MOSEI.

Model	MAE	Mul_Acc2	Mul_Acc5	Corr	F1_Score	CPU_Clock
TFN [11]	0.714	0.760	0.443	0.507	0.761	0.417
LMF [12]	0.729	0.761	0.436	0.520	0.760	0.412
MFN [9]	0.715	0.773	0.432	0.530	0.772	0.418
Gragh-MFN [10]	0.714	0.765	0.448	0.526	0.766	0.46
MARM [20]	0.708	0.772	0.449	0.530	0.773	0.363
LSTHM [20]-AttFusion	0.852	0.733	0.383	0.403	0.733	1.585
LSTHM [20]-Concat	0.861	0.704	0.383	0.383	0.721	1.6
LSTHM [20]-DFG	0.837	0.748	0.391	0.437	0.748	0.369
LSTHM [20]-Out_Product	0.905	0.722	0.383	0.405	0.723	0.715
MISA [41]	0.600	0.858	0.538	0.776	0.857	0.975
DISRFN (ours)	0.591	0.875	0.541	0.781	0.875	0.948

TABLE 4: Experiments of fusion methods.

Method	MAE (\downarrow)	Mul_Acc2 (\uparrow)	Mul_Acc5 (\uparrow)	Corr (\uparrow)	F1_Score (p/g) (\uparrow)
JDSN-AttFusion	0.924	0.791	0.378	0.687	0.782
JDSN-concat	0.839	0.814	0.443	0.724	0.813
JDSN-DFG	0.825	0.816	0.459	0.727	0.817
DISRFN (ours)	0.798	0.834	0.468	0.734	0.836

necessary. And it is also verified that the fusion network can dynamically fuse the multimodal data.

5.3. Ablation Study. The loss functions of various components discussed in Section 3.3 play an important role in the implementation of an improved joint domain separation network in Section 3.2. Therefore, the loss function is analyzed and discussed, and visualised and quantitative analysis is conducted based on ablation study.

5.3.1. Visual Presentation. An ablation experiment is designed in this section. The network is retrained after obtaining a zero setting of the loss weights ($\alpha, \beta, \lambda, \eta$) of other components except for the basic task loss L_{task} , and the best performance model parameters are saved. Moreover, to intuitively observe the effects of various loss functions on the

model results, the fusion representation of MOSI test samples is visualized by T-SNE, as shown in Figure 5.

As shown in Figure 5, the red spots represent positive emotions, and the blue ones represent negative emotions. When the distance between spots of the same color is shorter and the distance between spots of different colors is farther, the effect of semantic clustering and emotion analysis is better. The figure shows the T-SNE graph of the test data fusion representation, showing different distribution features under different loss function training. When all component losses exist, the model has the best semantic clustering effect. When the weight γ of the reconstruction loss L_{recon} is zero, it has the suboptimal clustering effect. When similarity loss L_{sim} does not exist, the clustering effect of the model is the most divergent. The impact of the loss L_{diff} and L_{trip} is between similarity loss and reconstruction loss. Furthermore, to explore the effect of each loss more

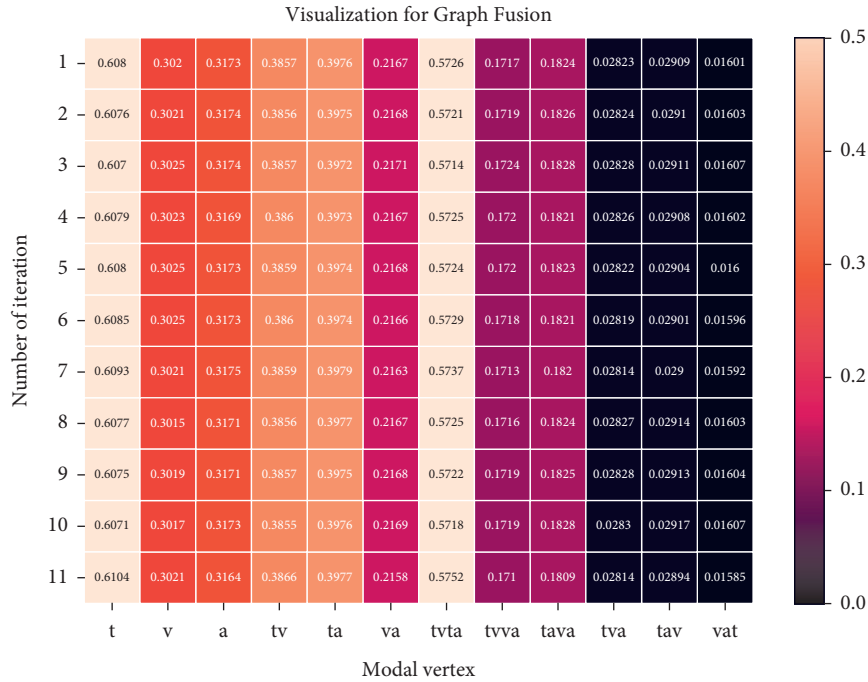
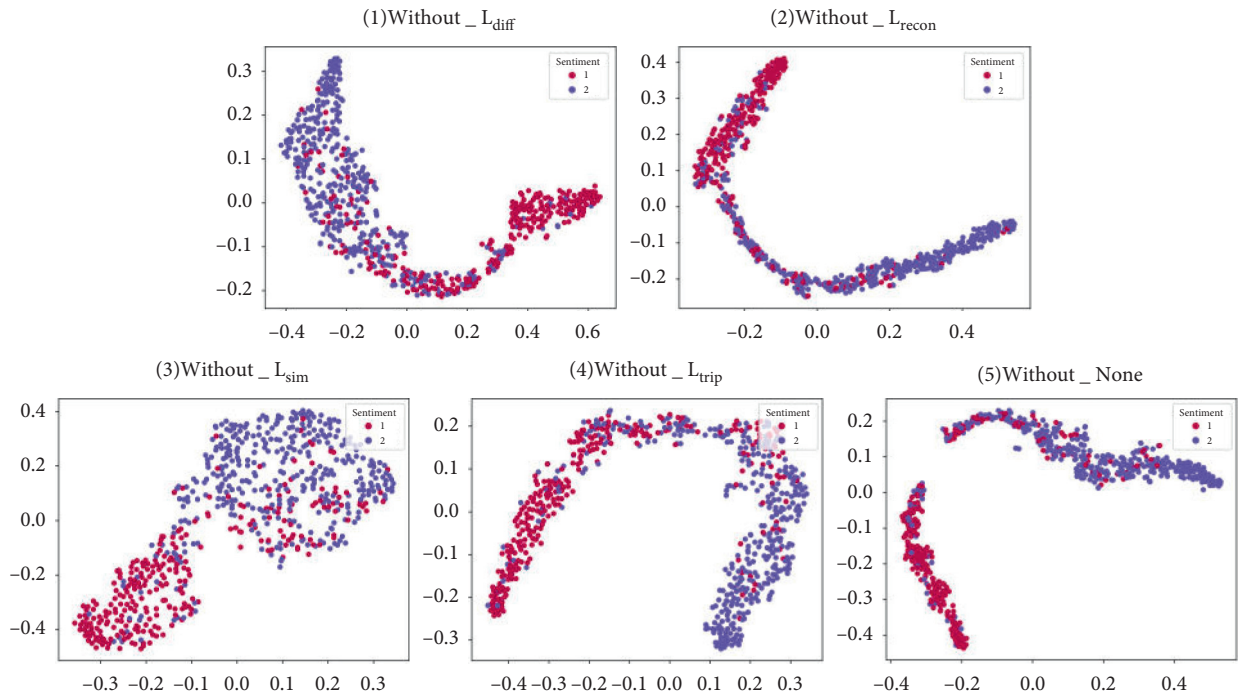


FIGURE 4: Visualization for Graph Fusion in MOSI sentiment analysis task.

FIGURE 5: Visualization of sentiment semantic distribution under different loss. Notes: (1) lack of loss function L_{diff} ; (2) lack of loss function L_{recon} ; (3) lack of loss function L_{sim} ; (4) lack of loss function L_{trip} ; (5) full configuration of loss function.

specifically, the evaluation indexes of the best model of each experiment are recorded in Table 5 for quantitative analysis.

5.3.2. Quantitative Analysis. As shown in Table 5, the model achieves the best performance when all losses are involved.

This finding indicates that each component loss is effective. The observation results show that the model is sensitive to L_{sim} and L_{diff} . It means that decomposing modes into independent space is conducive to the performance improvement of the model. The effect of cosine triplet-margin loss on the model is smaller than L_{sim} and L_{diff} . Because

TABLE 5: Experiments of ablation study.

Method	MAE	Mul_Acc2	Mul_Acc5	Corr	F1_Score
Without diff loss	0.868	0.811	0.404	0.728	0.816
Without sim loss	0.999	0.784	0.351	0.723	0.782
Without recon loss	0.833	0.817	0.464	0.711	0.816
Without CosineTriplet loss	0.857	0.799	0.469	0.705	0.798
ALL loss	0.798	0.834	0.468	0.734	0.836

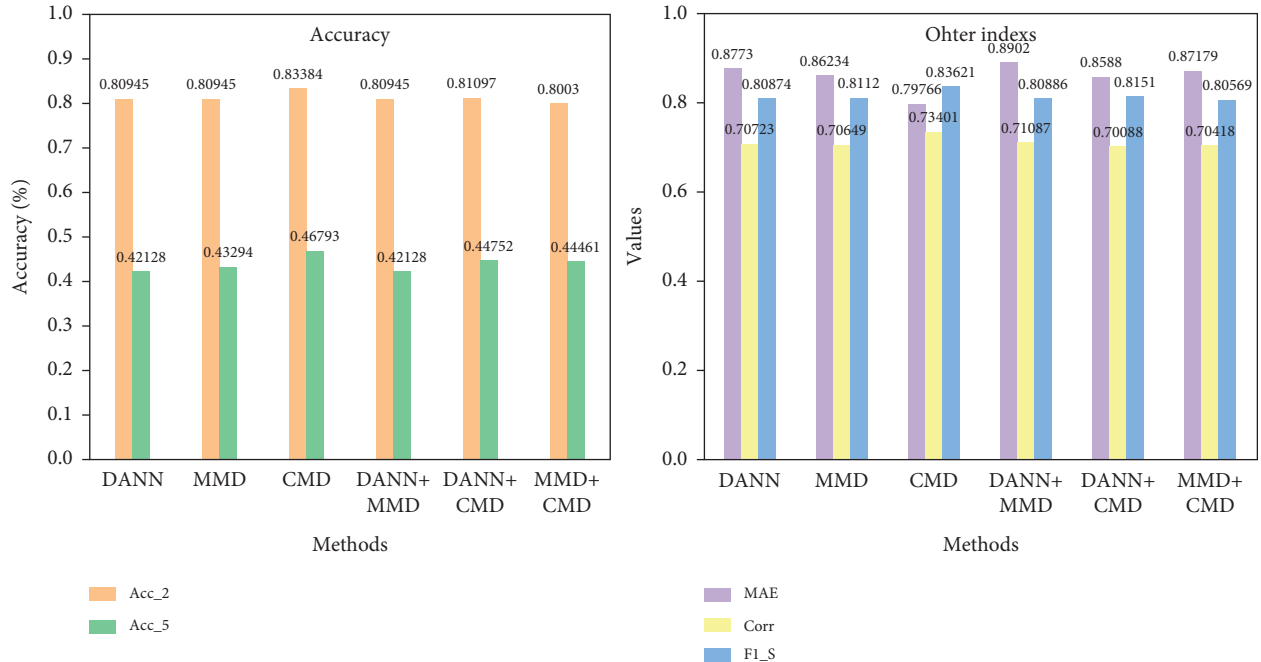


FIGURE 6: Visualization of performance comparison under different similarity loss.

semantic clustering effect is observed in the process of modal similarity feature acquisition. Therefore, the effect of this loss is weakened. In addition, the model is less dependent on reconstruction loss. The reason is that the trivial representation features of a specific encoder can be learned by L_{task} in the absence of reconstruction loss. The model is most sensitive to similarity loss; thus, the selection of similarity loss is very important. Therefore, an in-depth analysis is discussed in the following section.

5.4. Comparison of Similarity Measures. In this section, the selection of similarity loss function in 3.4.2 is discussed. For this reason, the following experiment is designed. Domain adversarial loss (DANN) [48], maximum mean square measure (MMD) [49], CMD, and their combinations are used for network training tests, as shown in Figure 6. The first three columns in the figure show that the performance of CMD in a single form is better than that of MMD and DANN in various indexes.

The reasons are summarised in the following points: (i) CMD can directly perform exact matching of the high-order moment without expensive distance and kernel matrix calculation; (ii) compared with CMD, DANN obtains modal similarity through minimax game using discriminator and

shared encoder. However, in adversarial training, additional parameters are added, and fluctuations may be encountered in training. Moreover, through the observation of joint form (the last three columns), the effect of similarity loss with CMD is better than that of the loss without CMD but worse than that of single CMD loss. This finding indicates that the increase in computation cost reduces the efficiency of network learning and further verifies the rationality of selecting CMD as similarity loss.

6. Conclusions

This paper studies multimodal emotion analysis. In the research, we have the following findings: (1) feature representation with more comprehensive information can reduce the burden of fusion network; (2) the redundant information of each mode can be used more effectively by jointing modality-invariance and modality-specificity representations of each mode; (3) simple dynamic fusion mechanism can obtain the interaction between modes more efficiently. Thus, this study puts forward a multimodal sentiment analysis framework consisting of two parts, namely, improved JDSN and HGFN. Firstly, modal invariant-specific joint representation of each mode is obtained through an improved JDSN module to effectively

utilize the complementary information amongst modes and reduce the heterogeneity gap between modes. Then, the joint representation of each mode is input to the HGFN for fusion to provide input for the prediction network. Moreover, a new combined loss function is designed to encourage the DISRFN model to learn the representation of expectation. Finally, the performance analysis experiment is carried out on MOSI and MOSEI data sets, obtaining acceptable results. In practice, the multimodal data usually have an unbalanced phenomenon, which will lead to the task bottleneck of the model. However, the study does not consider this issue. Therefore, we plan to study the problems of multimodal imbalance in the future.

Data Availability

The data used includes MOSI and MOSEI. The address of the MOSI dataset is correct. The MOSEI dataset address is as follows: http://immortal.multicomp.cs.cmu.edu/raw_data_sets/CMU_MOSEI.zip.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Changfan Zhang and Haonan Yang conceived and designed the experiments; Haonan Yang proposed the method; Jing He performed the experiments; Yifu Xu analyzed the data; Hongrun Chen prepared the original draft.

Acknowledgments

This work was supported by the Natural Science Foundation of China (U1934219, 52172403, and 62173137), Hunan Provincial Natural Science Foundation of China (2021JJ50001 and 2021JJ30217), and Project of Hunan Provincial Department of Education (19A137).

References

- [1] Q. Li, D. Gkoumas, C. Lioma, and M. Melucci, "Quantum-inspired multimodal fusion for video sentiment analysis," *Information Fusion*, vol. 65, pp. 58–71, 2021.
- [2] K. Huang, W. Zhou, and M. Fang, "Deep multimodal fusion autoencoder for saliency prediction of RGB-D images," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–10, 2021.
- [3] J. Zhou, M. Ye, J. Ding, S. Mao, and H. J. Zhang, "Rapid and robust traffic accident detection based on orientation map," *Optical Engineering*, vol. 51, no. 11, Article ID 117201, 2012.
- [4] J. Yu, J. Jiang, and R. Xia, "Entity-sensitive attention and fusion network for entity-level multimodal sentiment classification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 429–439, 2020.
- [5] S. Mao, M. Ye, X. Li, F. Pang, and J. Zhou, "Rapid vehicle logo region detection based on information theory," *Computers & Electrical Engineering*, vol. 39, no. 3, pp. 863–872, 2013.
- [6] S. Mao, H. Wu, and M. Lu, "Multiple 3D marker localization and tracking system in image-guided radiotherapy," *International Journal of Robotics and Automation*, vol. 32, no. 5, pp. 517–523, 2017.
- [7] Y. Zhang, D. Song, X. Li et al., "A Quantum-Like multimodal network framework for modeling interaction dynamics in multiparty conversational sentiment analysis," *Information Fusion*, vol. 62, pp. 14–31, 2020.
- [8] S. Agethen and W. H. Hsu, "Deep multi-kernel convolutional LSTM networks and an attention-based mechanism for videos," *IEEE Transactions on Multimedia*, vol. 22, no. 3, pp. 819–829, 2020.
- [9] A. Zadeh, P. P. Liang, N. Mazumder, S. Poria, E. Cambria, and L.-P. Morency, "Memory fusion network for multiview sequential learning," in *proceedings of the thirty-second AAAI conference on artificial intelligence (AAAI-2018)*, vol. 32, no. 1, pp. 5634–5641, New Orleans, Louisiana, USA, 2018.
- [10] A. Zadeh, P. P. Liang, J. Vanbriesen, S. Poria, E. Cambria, and L. Morency, "multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL-2018)*, pp. 2236–2246, Melbourne, Australia, 2018.
- [11] A. Zadeh, M. Chen, S. Poria, E. Cambria, and L.-P. Morency, "Tensor fusion network for multimodal sentiment analysis," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, vol. 1, pp. 1103–1114, Copenhagen, Denmark, 2017.
- [12] Z. Liu, Y. Shen, V. B. Lakshminarasimhan, P. P. Liang, A. Zadeh, and L.-P. Morency, "Efficient low-rank multimodal fusion with modality-specific factors," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 2247–2256, Melbourne, Australia, 2018.
- [13] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, "Domain separation networks," in *Proceedings of the 30th conference on neural information processing systems (NIPS-2016)*, vol. 3, pp. 343–351, Barcelona, Spain, 2016.
- [14] R. F. Silva, S. M. Plis, T. Adali, M. S. Pattichis, and V. D. Calhoun, "Multidataset independent subspace analysis with application to multimodal fusion," *IEEE Transactions on Image Processing*, vol. 30, pp. 588–602, 2021.
- [15] S. Mai, H. Hu, and S. Xing, "Modality to Modality Translation: An Adversarial Representation Learning and Graph Fusion Network for Multimodal Fusion," in *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 164–172, New York, NY, USA, 2020.
- [16] J. Kim, T. Kim, S. Kim, and C. Yoo, "Edge-labeling graph neural network for few-shot learning," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019.
- [17] S. J. Mai, H. F. Hu, and S. L. X. Divide, "Conquer and combine: hierarchical feature fusion network with local and global perspectives for multimodal affective computing," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 481–492, Florence Italy, 2019.
- [18] S. Mai, S. Xing, and H. Hu, "Locally confined modality fusion network with a global perspective for multimodal human affective computing," *IEEE Transactions on Multimedia*, vol. 22, no. 1, pp. 122–137, 2020.
- [19] P. P. Liang, Z. Liu, A. Zadeh, and L. P. Morency, "multimodal language analysis with recurrent multistage fusion," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, vol. 1, pp. 150–161, Brussels, Belgium, 2018.

- [20] A. Zadeh, P. P. Liang, S. Poria, P. Vij, E. Cambria, and L.-P. Morency, "Multi-attention recurrent network for human communication comprehension," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, pp. 5642–5649, New Orleans, Louisiana, USA, 2018.
- [21] M. Hou, J. Tang, J. Zhang, W. Kong, and Q. Zhao, "Deep multimodal multilinear fusion with high-order polynomial pooling," in *Proceedings of the advances in neural information processing systems*, vol. 32, pp. 1–10, Vancouver, Canada, 2019.
- [22] S. Mai, H. Hu, J. Xu, and S. Xing, "Multi-fusion residual memory network for multimodal human sentiment comprehension," *IEEE Transactions On Affective Computing*, p. 1, 2020.
- [23] Y.-J. Zhang and Z.-H. Ling, "Extracting and predicting word-level style variations for speech synthesis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1582–1593, 2021.
- [24] Y. Wang, Y. Shen, Z. Liu, P. P. Liang, A. Zadeh, and L.-P. Morency, "Words can shift: dynamically adjusting word representations using nonverbal behaviors," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 7216–7223, Honolulu, Hawaii, 2019.
- [25] M. H. Chen, S. Wang, P. P. Liang, T. Baltrušaitis, A. Zadeh, and L.-P. Morency, "Multimodal sentiment analysis with word-level fusion and reinforcement learning," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI-17)*, pp. 163–171, Glasgow, Scotland, 2017.
- [26] X. Shu and G. Zhao, "Scalable multilabel canonical correlation analysis for cross-modal retrieval," *Pattern Recognition*, vol. 115, Article ID 107905, 2021.
- [27] Y. Kaloga, P. Borgnat, S. P. Chepuri, P. Abry, and A. Habrard, "Variational graph autoencoders for multiview canonical correlation analysis," *Signal Processing*, vol. 104, Article ID 108182, 2021.
- [28] S. Verma, J. W. Wang, Z. F. Ge et al., "Deep-HOSeq: Deep Higher Order Sequence Fusion for Multimodal Sentiment Analysis," in *Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM)*, Sorrento, Italy, 2020.
- [29] H. Pham, P. P. Liang, T. Manzini, L.-P. Morency, and B. Póczos, "Found in translation: learning robust joint representations by cyclic translations between modalities," in *Proceedings of the thirty-third AAAI conference on artificial intelligence (AAAI-19)*, vol. 33, no. 01, pp. 6892–6899, Honolulu, Hawaii, USA, 2019.
- [30] H. Qiang, Y. Wan, L. Xiang, and X. Meng, "Deep semantic similarity adversarial hashing for cross-modal retrieval," *Neurocomputing*, vol. 400, pp. 24–33, 2020.
- [31] F. Wu, X.-Y. Jing, Z. Wu et al., "Modality-specific and shared generative adversarial network for cross-modal retrieval," *Pattern Recognition*, vol. 104, Article ID 107335, 2020.
- [32] Y. H. H. Tsai, P. P. Liang, A. Zadeh, L. Morency, and R. Salakhutdinov, "Learning factorized multimodal representations," in *Proceedings of the International Conference on Learning Representations (ICLR-2019)*, New Orleans, Louisiana, USA, 2019.
- [33] P. P. Liang, Y. C. Lim, Y. H. Tsai, R. Salakhutdinov, and L. Morency, "Strong and simple baselines for multimodal utterance embeddings," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, vol. 1, pp. 2599–2609, Minneapolis, Minnesota, 2019.
- [34] D. Wang, Q. Wang, L. He, X. Gao, and Y. Tian, "Joint and individual matrix factorization hashing for large-scale cross-modal retrieval," *Pattern Recognition*, vol. 107, Article ID 107479, 2020.
- [35] Y. Fang, Y. Ren, and J. H. Park, "Semantic-enhanced discrete matrix factorization hashing for heterogeneous modal matching," *Knowledge-Based Systems*, vol. 192, Article ID 105381, 2020.
- [36] J. C. Caicedo, J. Benabdallah, F. A. González, and O. Nasraoui, "Multimodal representation, indexing, automated annotation and retrieval of image collections via non-negative matrix factorization," *Neurocomputing*, vol. 76, no. 1, pp. 50–60, 2012.
- [37] Y. Wu, Y. Zhao, X. Lu et al., "Modeling incongruity between modalities for multimodal sarcasm detection," *IEEE Multimedia*, vol. 28, pp. 86–95, 2021.
- [38] F. Chen, Z. Luo, and Y. Xu, "Complementary Fusion of Multi-Features and Multi-Modalities in Sentiment Analysis," in *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*, New York, NY, USA, 2020.
- [39] Z. Sun, P. Sarma, W. Sethares, and Y. Liang, "Learning relationships between text, audio, and video via deep canonical correlation for multimodal language analysis the thirty-fourth AAAI conference on artificial intelligence (AAAI-20)," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, pp. 8992–8999, New York, NY, USA, 2020.
- [40] W. Guo, J. Wang, and S. Wang, "Deep multimodal representation learning: a survey," *IEEE Access*, vol. 7, Article ID 63373, 2019.
- [41] D. Hazarika, R. Zimmermann, and S. Poria, "MISA: Modality-invariant and -specific representations for multimodal sentiment analysis," in *Proceedings of the 28th ACM International Conference on Multimedia (ACM MM)*, pp. 1122–1131, Seattle, USA, 2020.
- [42] Y. Aytar, L. Castrejon, C. Vondrick, H. Pirsiavash, and A. Torralba, "Cross-modal scene networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 10, pp. 2303–2314, 2018.
- [43] W. Zelling, T. Grubinger, E. Lughofer, T. Natschlager, and S. S. Platz, "CMD for Domain-Invariant Representation learning," in *Proceedings of the 5th International Conference on Learning Representations (ICLR-2017)*, Toulon, France, 2017.
- [44] W. Gu, X. Y. Gu, J. Z. Gu, B. Li, Z. Xiong, and W. Wang, "Adversary guided asymmetric hashing for cross-modal retrieval," in *Proceedings of the 2019 International Conference on Multimedia Retrieval (ICMR-2019)*, pp. 159–167, Ottawa, ON, Canada, 2019.
- [45] J. He, S. Mai, and H. Hu, "A unimodal reinforced transformer with time squeeze fusion for multimodal sentiment analysis," *IEEE Signal Processing Letters*, vol. 28, pp. 992–996, 2021.
- [46] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, "COVAREP-A collaborative voice analysis repository for speech technologies," in *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 960–964, Florence, Italy, 2014.
- [47] R. Aharoni and Y. Goldberg, "Unsupervised domain clusters in pretrained language models," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL-2020)*, vol. 1, pp. 7747–7763, Seattle, Washington, USA, 2020.

- [48] H. Tang and K. Jia, "Vicinal and categorical domain adaptation," *Pattern Recognition*, vol. 115, Article ID 107907, 2021.
- [49] J. Pomponi, S. Scardapane, and A. Uncini, "Bayesian neural networks with maximum mean discrepancy regularization," *Neurocomputing*, vol. 453, pp. 428–437, 2021.