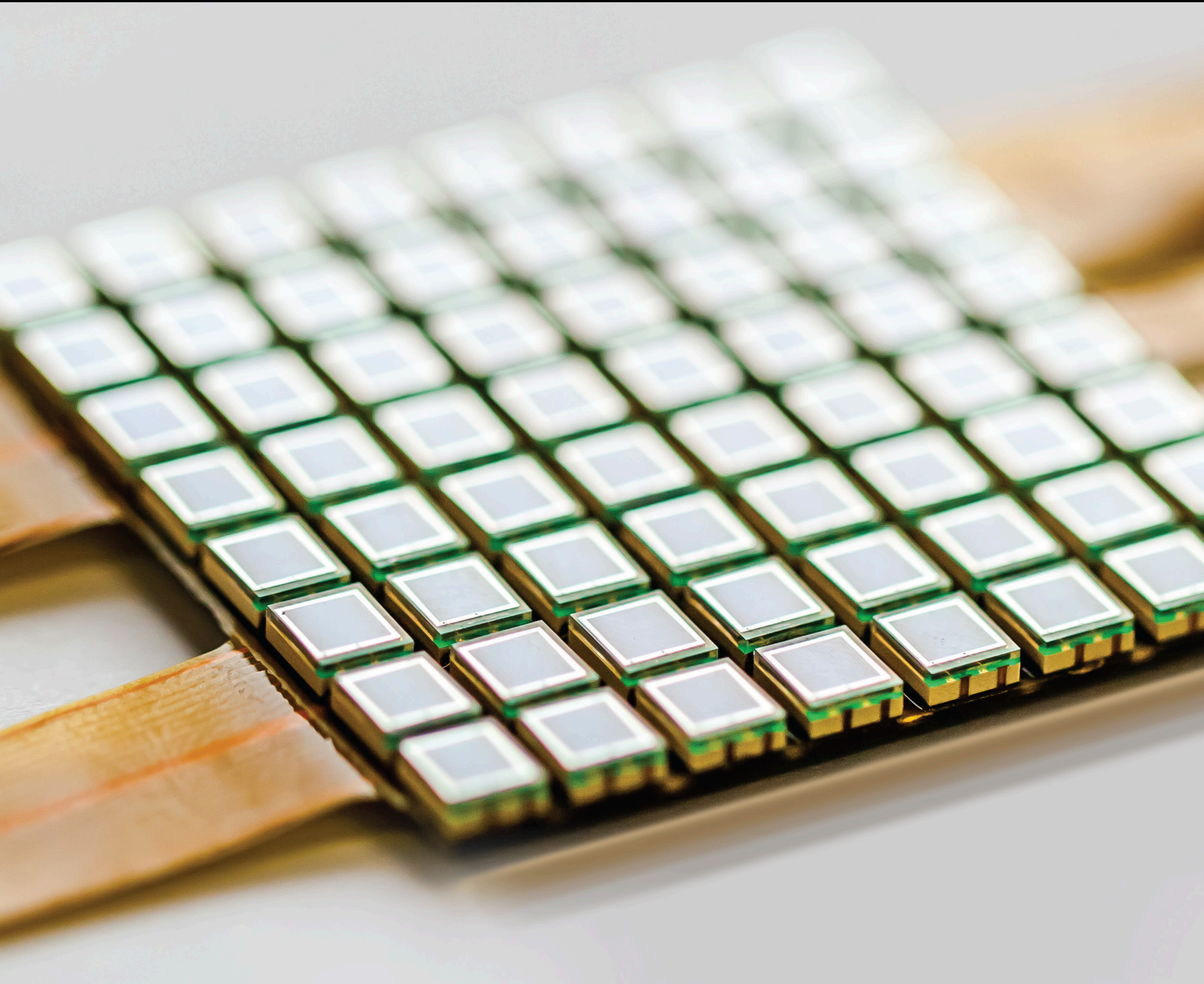


Integration of Sensors in Control and Automation Systems 2020

Lead Guest Editor: Rafael Morales

Guest Editors: Antonio Fernández-Caballero, José Andrés Somolinos, and Hebertt Sira-Ramírez





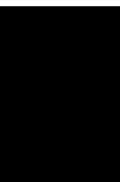
Integration of Sensors in Control and Automation Systems 2020

Journal of Sensors

Integration of Sensors in Control and Automation Systems 2020

Lead Guest Editor: Rafael Morales

Guest Editors: Antonio Fernández-Caballero, José Andrés Somolinos, and Hebertt Sira-Ramírez



Copyright © 2022 Hindawi Limited. All rights reserved.

This is a special issue published in "Journal of Sensors." 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

Harith Ahmad, Malaysia

Editorial Board

Ghufran Ahmed, Pakistan
Manuel Aleixandre, Spain
Bruno Andò, Italy
Constantin Apetrei, Romania
Fernando Benito-Lopez, Spain
Romeo Bernini, Italy
Shekhar Bhansali, USA
Matthew Brodie, Australia
Belén Calvo, Spain
Stefania Campopiano, Italy
Binghua Cao, China
Domenico Caputo, Italy
Sara Casciati, Italy
Gabriele Cazzulani, Italy
Chi Chiu Chan, Singapore
Edmon Chehura, United Kingdom
Marvin H Cheng, USA
Mario Collotta, Italy
Marco Consales, Italy
Jesus Corres, Spain
Andrea Cusano, Italy
Dzung Dao, Australia
Egidio De Benedetto, Italy
Luca De Stefano, Italy
Manel del Valle, Spain
Francesco Dell'Olio, Italy
Franz L. Dickert, Austria
Giovanni Diraco, Italy
Nicola Donato, Italy
Mauro Epifani, Italy
Congbin Fan, China
Vittorio Ferrari, Italy
Luca Francioso, Italy
Bin Gao, China
Libo Gao, China
Manel Gasulla, Spain
Carmine Granata, Italy
Banshi D. Gupta, India
Mohammad Haider, USA
Agustin Herrera-May, Mexico
María del Carmen Horrillo, Spain
Evangelos Hristoforou, Greece
Shahid Hussain, China
Syed K. Islam, USA

Stephen James, United Kingdom
Bruno C. Janegitz, Brazil
Hai-Feng Ji, USA
Sang Sub Kim, Republic of Korea
Antonio Lazaro, Spain
Chengkuo Lee, Singapore
Yuan Li, China
Chenzong Li, USA
Rosalba Liguori, Italy
Sangsoon Lim, Republic of Korea
Duo Lin, China
Eduard Llobet, Spain
Jaime Lloret, Spain
Yu-Lung Lo, Taiwan
Jesús Lozano, Spain
Oleg Lupan, Moldova
Frederick Mailly, France
Pawel Malinowski, Poland
Vincenzo Marletta, Italy
Carlos Marques, Portugal
Eugenio Martinelli, Italy
Antonio Martinez-Olmos, Spain
Giuseppe Maruccio, Italy
Yasuko Y. Maruo, Japan
Fanli Meng, China
Carlos Michel, Mexico
Stephen. J. Mihailov, Canada
Heinz C. Neitzert, Italy
Sing Kiong Nguang, New Zealand
Calogero M. Oddo, Italy
Marimuthu Palaniswami, Australia
Alberto J. Palma, Spain
Davide Palumbo, Italy
Roberto Paolesse, Italy
Giovanni Pau, Italy
Giorgio Pennazza, Italy
Michele Penza, Italy
Salvatore Pirozzi, Italy
Antonina Pirrotta, Italy
Stelios M. Potirakis, Greece
Biswajeet Pradhan, Malaysia
Giuseppe Quero, Italy
Valerie Renaudin, France
Armando Ricciardi, Italy



Christos Riziotis, Greece
Maria Luz Rodriguez-Mendez, Spain
Jerome Rossignol, France
Carlos Ruiz, Spain
Ylias Sabri, Australia
José P. Santos, Spain
Sina Sareh, United Kingdom
Isabel Sayago, Spain
Andreas Schütze, Germany
Praveen K. Sekhar, USA
Sandra Sendra, Spain
Pietro Siciliano, Italy
Vincenzo Spagnolo, Italy
Sachin K. Srivastava, India
Grigore Stamatescu, Romania
Stefano Stassi, Italy
Vincenzo Stornelli, Italy
Salvatore Surdo, Italy
Yunchao Tang, China
Roshan Thotagamuge, Sri Lanka
Guiyun Tian, United Kingdom
Vijay Tomer, USA
Abdellah Touhafi, Belgium
Hoang Vinh Tran, Vietnam
Aitor Urrutia, Spain
Hana Vaisocherova - Lisalova, Czech
Republic
Everardo Vargas-Rodriguez, Mexico
Xavier Vilanova, Spain
Luca Vollero, Italy
Tomasz Wandowski, Poland
He Wen, China
Qihao Weng, USA
Qiang Wu, United Kingdom
Penghai Wu, China
Chen Yang, China
Aijun Yin, China
Chouki Zerrouki, France

Contents


Integration of Sensors in Control and Automation Systems 2020

Rafael Morales , Antonio Fernández-Caballero , José A. Somolinos, and Hebertt Sira-Ramírez 
Editorial (3 pages), Article ID 9765679, Volume 2022 (2022)

An Industrial B-Mode Phased Array Ultrasonic Imaging Reconstruction Algorithm Based on FRI Sampling

Guangzhi Dai , Zhiyong He , and Song Lin
Research Article (11 pages), Article ID 8844741, Volume 2021 (2021)


A Unified Calibration Paradigm for a Better Cuffless Blood Pressure Estimation with Modes of Elastic Tube and Vascular Elasticity

Jiang Shao, Ping Shi , and Sijung Hu
Research Article (12 pages), Article ID 8868083, Volume 2021 (2021)



-th Order Sensor Output to Control -DoF Serial Robot Arms

Sergio Alvarez-Rodríguez  and Francisco Gerardo Peña Lecona 
Research Article (14 pages), Article ID 8884282, Volume 2021 (2021)


Research on Innovative Trim Method for Tiltrotor Aircraft Take-Off Based on Genetic Algorithm

Xueyun Wang , Jiyang Chen, Qian Zhang , Jingjuan Zhang, and Hao Cong
Research Article (13 pages), Article ID 8876867, Volume 2020 (2020)


A Step Length Estimation Model of Coefficient Self-Determined Based on Peak-Valley Detection

Wenxia Lu , Fei Wu , Hai Zhu, and Yujin Zhang
Research Article (14 pages), Article ID 8818130, Volume 2020 (2020)

Design of a Secure Wireless Home Automation System with an Open Home Automation Bus (OpenHAB 2) Framework

Robert A. Sowah , Dale E. Boahene, Dalton C. Owoh, Rexford Addo, Godfrey A. Mills, Wiafe Owusu-Banahene, Gifty Buah, and Baffour Sarkodie-Mensah
Research Article (22 pages), Article ID 8868602, Volume 2020 (2020)

Application of Artificial Intelligence Techniques in Predicting the Lost Circulation Zones Using Drilling Sensors

Abdulmalek Ahmed, Salaheldin Elkatatny , Abdulwahab Ali, Mahmoud Abughaban, and Abdulazeez Abdulraheem
Research Article (18 pages), Article ID 8851065, Volume 2020 (2020)

Editorial

Integration of Sensors in Control and Automation Systems 2020

Rafael Morales ¹, **Antonio Fernández-Caballero** ¹, **José A. Somolinos**,²
and Hebertt Sira-Ramírez ³

¹Escuela Técnica Superior de Ingenieros Industriales de Albacete, Universidad de Castilla-La Mancha, 02071 Albacete, Spain

²Escuela Técnica Superior de Ingenieros Navales, Universidad Politécnica de Madrid, Madrid, Spain

³Center for Research and Advanced Studies, National Polytechnic Institute, Mexico City, Mexico

Correspondence should be addressed to Rafael Morales; rafael.morales@uclm.es

Received 30 November 2021; Accepted 30 November 2021; Published 25 January 2022

Copyright © 2022 Rafael Morales 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.

1. Introduction

The integration of sensors in control and automation systems has received a great deal of attention from a considerable number of researchers and the industrial community in the last years. Emphasis is placed on the importance of creating improvements in control and automation systems to meet the challenges of developing and refining new applications. These systems have to integrate a variety of sensory information and human knowledge for the sake of efficiently carrying out tasks with or without human intervention. In fact, the integration of sensors into intelligent devices and systems has increased the capacity to measure, analyze, and aggregate data at a local level. Autonomous and connected sensors are able to selectively sample and measure many physical properties. Built on the increasing capabilities of fixed-access and wireless networks, smart sensor developments allow the collection of raw data, which are processed into information and conveyed via a network connection.

The concept of sensor integration is close to the sensor fusion term, which is defined as “the art of processing data from multiple sensors with an aim to replicate a physical environment or induce intelligence to control a phenomenon with increased precision and reliability.” Sensor fusion or integration is evolving rapidly as the basis of robust control systems that make sense of imperfect input despite the environment in which it operates. Data from multiple sensors are fused to increase response and accuracy, delivering control systems that until recently could only be theorized,

drawing on techniques like artificial intelligence, pattern recognition, digital signal processing, and statistical estimation. Moreover, recent advances in sensor technology and processing techniques, combined with improved hardware, make real-time data fusion possible.

This special issue was aimed at exhibiting the latest research achievements, findings, and ideas in the integration of sensors in control and automation systems. The topics faced in this special issue were the following:

- (i) Sensor systems for control and automation: sensors and sensor networks, intelligent sensors, sensor uncertainty for fault-tolerant control, distributed and multimodality sensor network for control and automation, and so on
- (ii) Control: adaptive control, robust control, active disturbance rejection control, complex systems, identification and estimation, nonlinear systems, intelligent systems, sensor networks, delay systems, precision motion control, control applications, and so on
- (iii) Automation: man-machine interactions, process automation, network-based systems, intelligent automation, planning, scheduling and coordination, and so on
- (iv) Robotics: modelling and identification, mobile robotics, mobile sensor networks, perception systems,

visual servoing, robot sensing and data fusion, and so on

- (v) Process-based control: sensor development, system design, and control development
- (vi) Control and automation systems: fault detection and isolation, sensing and data fusion, flight control and surveillance systems, rescue and field robotics, guidance control systems, industry, military, space and underwater applications, linear and nonlinear control systems, signal and image processing, and so on
- (vii) Industrial informatics: embedded systems for monitoring and controlling

2. The Papers

A total of 20 papers were submitted to this special issue. After peer review, finally, 7 were accepted and published, covering a wide range of the topics proposed in the call for papers.

J. Shao et al. proposed a unified calibration paradigm for a better cuffless blood pressure (BP) estimation with modes of elastic tube (ET) and vascular elasticity (VE). The study was aimed at evaluating the performance of VE and ET models by means of an advanced point-to-point pairing calibration. With the study, a cost-effective cuffless BP monitoring approach could be emerged with an easy and durable personalized calibration. Such an approach could be anticipated to be a better choice when considering the practicality of long-term and continuous BP monitoring with both modes of elastic tube and vascular elasticity. Besides these, the study was proofed evidence about the sensitivity of BP estimation along with these models and their initial calibration methods.

G. Dai et al. proposed an industrial B-mode phased array ultrasonic imaging reconstruction algorithm based on finite rate of innovation (FRI). The new FRI sampling model had the advantages of its good stability, simple circuit, and implementation. Additionally, the B-mode phased array ultrasonic imaging algorithm was proposed based on the FRI sampling model and the mathematical model characteristics of B-mode phased array ultrasonic imaging. The simulation results indicated that the sampling point required by the proposed FRI sampling model is 0.1% of the traditional mode of B-mode phased array ultrasonic imaging, and the sampling frequency of the proposed ultrasonic imaging algorithm is 0.0077% of that of the traditional B-mode ultrasonic imaging.

S. Álvarez-Rodríguez and F.G. Peña Lecona investigated the performance of a κ -degree of freedom serial robot arm with dynamical inclusion of linear n -order sensors, showing that robot's properties with linear n -order sensor inclusion were invariant with respect to robot's theoretical dynamics, provided that the solutions of the considered linear n -order sensors exist and are unique. The proposed methodology was demonstrated by a formal proof, and additionally, the effectiveness of the proposed method was validated by

means of the implementation of a trajectory tracking control problem.

X. Wang et al. developed an innovative trim method for tiltrotor aircraft take-off based on a genetic algorithm. Firstly, the genetic algorithm, which possesses strong capability in searching global optimum, was adopted to identify a coarse solution. Secondly, the coarse solution of the trim is further refined by the Levenberg-Marquardt method for precise local optimum. In addition, the innovative trim method was applied to a tiltrotor aircraft's flight control in the transition process of incline take-off. The limitation of trajectory was discussed, and the tilt corridor was constructed. Finally, the incline take-off simulations were conducted, and the effectiveness of the proposed trim method was verified through good match with the designed reference trajectory.

W. Lu et al. proposed a peak-valley detection method that detects the pair of peak-valley to overcome the problem of overcounting. Based on the fuzzy logic algorithm, a fuzzy controller was defined to make the constant coefficient in the Weinberg nonlinear step length estimation model to be adjusted adaptively to each detected step, which was suitable for different kinds of people walking at various velocities. It was also possible to estimate pedestrian walking distance accurately by accumulating every estimated step length.

R.A. Sowah et al. designed and developed a cost-effective, secure home automation using the OpenHAB 2 framework with capability for device programming and customizations. Additionally, they develop mobile and web applications for energy management and switching of connected home devices and interactive visual interface for home automation. It leveraged the developed hardware and software modules to provide optimal energy management for the home. They also implemented additional security layers of user authentication and authorization while keeping the overall cost of implementation low and maintaining the ease of deployment for everyday home use. Per our server configuration, the OpenHAB communication through the Internet is made through the JSON Web Token authentication procedure. This process makes it difficult for user identity to be hijacked by a malicious attacker. This approach proved to be more secure than the default OpenHAB server configuration. Consequently, we leveraged on the open-source OpenHAB REST API to develop a mobile or web application that is flexible and easily adaptable for traditional home use.

Finally, A. Ahmed et al. studied the application of artificial intelligence techniques in predicting the lost circulation zones using drilling sensors. In particular, they evaluated three AI techniques to predict the lost circulation zones based only on six mechanical surface drilling parameters. These techniques were functional networks, artificial neural networks, and fuzzy logic. On the other hand, the six parameters were real-time measurements of flow pump, rate of penetration, string rotary speed, standpipe pressure, drilling torque, and weight on bit. In the experiments developed, more than 4500 real-field data points from three wells were used in the evaluation.

Acknowledgments

We would like to thank all the authors for their excellent contributions, and also the reviewers for their valuable help. We would also like to thank all members of the editorial board for approving this special issue.

Rafael Morales
Antonio Fernández-Caballero
José A. Somolinos
Hebertt Sira-Ramírez

Research Article

An Industrial B-Mode Phased Array Ultrasonic Imaging Reconstruction Algorithm Based on FRI Sampling

Guangzhi Dai ¹, Zhiyong He ², and Song Lin³

¹School of Artificial Intelligence, Shenzhen Polytechnic, Shenzhen 518055, China

²School of Mechanical and Electric Engineering, Soochow University, Suzhou 215021, China

³School of Mechatronics Engineering, Harbin Institute of Technology, Harbin 150001, China

Correspondence should be addressed to Zhiyong He; he-zhiyong@139.com

Received 18 March 2020; Accepted 9 June 2021; Published 17 September 2021

Academic Editor: Rafael Morales

Copyright © 2021 Guangzhi Dai 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.

Firstly, a novel FRI sampling model has been proposed according to the characteristics of ultrasonic signals. The model has the advantages such as good stability, strong antinoise ability, simple circuit implementation, and fewer preconditions, compared to the traditional methods. Then, in order to verify the validity of the sampling model, the method is applied to B-type ultrasonic imaging, and a B-type phased array ultrasonic imaging algorithm based on FRI sampling model is proposed. Finally, the algorithm simulation experiment is designed, and the results show that the sampling point required by the proposed FRI sampling model is only 0.1% of the traditional B-type phased array ultrasonic imaging method, and the sampling frequency of the proposed ultrasonic imaging algorithm is only 0.0077% of the traditional B-type ultrasonic imaging method. Additionally, the experiment result indicates that this algorithm is more applicable to phased array ultrasonic imaging than the SOS filter is.

1. Introduction

With the development of ultrasonic testing technology, the number of transducers and the amount of recorded data have been increasing rapidly. Theories proposed in recent years, such as CS (compressed sensing), can break through the limitation of traditional sampling theories and obtain an accurate reconstruction of signals by sampling of sparse signals with lower frequencies. Such sampling method is called sub-Nyquist sampling or undersampling. An undersampling method develops fast and has made great progress in many practical applications: such as the single-pixel camera developed by Rice University, MRI RF pulse device and coded aperture camera developed by MIT, DNA microarray sensor developed by Illinois State University, and CS filter and chaotic device developed by Chinese Academy of Sciences. However, in the field of ultrasonic imaging, the applications of the undersampling have been rarely reported.

At present, researches on ultrasonic imaging technology based on the undersampling theory can be summarized into two categories:

- (1) Researches focusing on the improvement of the speed and quality of imaging, where two methods are developed to improve the imaging resolution: one of them changes the direction of the transducer, such as sound velocity focusing method; the other is postprocessing data, such as filtering and deconvolution methods. Due to the limitation of manufacturing technology and procedures, it is difficult for the transducers to get smaller focuses, which usually produce huge scanning data and therefore affect the postprocessing of the data. In the postprocessing of the data, the superresolution technology becomes a feasible and promising method [1–3]. However, the high resolution image of classic superresolution algorithm is often constructed from the low-resolution image; due to the deficiency of inherent data acquisition conditions and the problem of noise accumulation, it is hard to implement the algorithm. Therefore, a new theory is required to break through the above difficulties, so that the above problems can be solved. The appearance of the CS theory has laid a theoretical

foundation for rapid high-resolution ultrasonic imaging, attracting a lot of attentions from many researchers, such as Friboulet et al. [4–17] and Eldar et al. [18–34].

- (2) The researches focusing on a reduction of the volume and energy-consumption of imaging device: these researches are mainly concentrated on two aspects, one is the undersampling imaging technology based on the CS theory, proposed by Friboulet et al.; the other is the undersampling imaging technology based on FRI sampling, proposed by Eldar et al.

A fast imaging method based on undersampling theory is a research hotspot in the field of ultrasound imaging in recent years. Friboulet et al. first introduced CS theory into the field of medical ultrasound imaging in 2010 [4]. Using wavelet transform basis function and wave atom basis function as the sparse representation basis of ultrasonic image, the sparse imaging of ultrasonic image is realized. Subsequently, Friboulet et al. conducted in-depth research on CS in the field of medical ultrasound imaging [5–10], mainly carried out sparse analysis of medical ultrasound images in K space, and discussed in detail the reconstruction performance of different recovery algorithms based on CS for ultrasonic echo signal and extended its application to three-dimensional ultrasound imaging. In 2013, Friboulet et al. introduced the distributed CS theory [11, 12] into medical ultrasound imaging. They sparsely represent the echo signals of each transducer array on the same orthogonal basis of Fourier transform. Using the correlation between signals, joint sparse model and joint data recovery algorithm achieve CS imaging [13]. In addition, Eldar et al. first applied FRI theory to ultrasound in 2011 and proposed a sparse sampling structure for ultrasound imaging [18–34]. The research is based on the finite rate of innovation (FRI) [35] sampling signal model proposed by Vetterli et al. FRI uses Gaussian sampling kernel to sample, which overcomes the limitation of Shannon sampling width. However, this structure is unstable, and the signal-to-noise ratio is not ideal. Therefore, Tur et al. used FRI principle to construct measurement matrix and realized ultrasonic data imaging under single channel sampling framework [23]. Later, the research team conducted research on multichannel ultrasonic data sampling and frequency domain sparse sampling [24] and even explored the superresolution ultrasonic reconstruction algorithm based on CS [29]. Guangming et al. of Xidian University [36, 37] and Chinese Academy of Sciences [38] also developed various applications of CS in ultrasonic imaging. To sum up, FRI theory is still in the stage of theoretical research, and FRI sampling based on ultrasonic signal is still in the stage of preliminary exploration. Therefore, there are still some key problems to be studied in theory and application.

In this paper, an under sampling method for ultrasonic testing signal is proposed, which is a multichannel indirect FRI sampling method based on ultrasonic signal. In each channel, the original signal is multiplied by the carrier to generate a new analog signal, and then, the new signal is integrated and sampled to obtain a group of linear transfor-

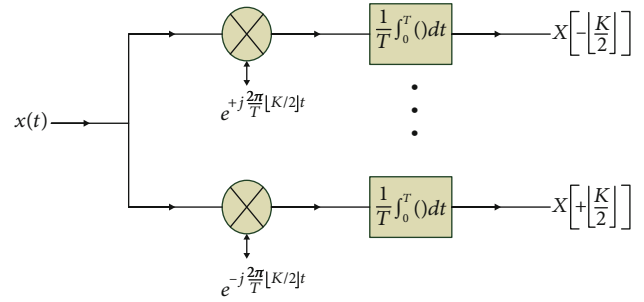


FIGURE 1: Multichannel direct sampling model.

mations about the Fourier coefficients of the original signal. Then, the Fourier coefficients of the original signal are obtained through linear transformation, and the signal is reconstructed by a spectral method. This scheme not only has the advantages of high stability and strong antinoise ability but also has the advantages of simple circuit structure, less need for prior knowledge, and much lower sampling frequency than traditional sampling methods. In order to verify the effectiveness of the proposed FRI sampling model, this method is applied to B-mode ultrasonic imaging, and a B-mode phased array ultrasonic imaging algorithm based on the FRI sampling model is proposed. The algorithm is based on the multichannel indirect FRI sampling model of ultrasonic signal and combined with the line scan mathematical model of B-mode ultrasonic imaging, so as to derive the B-mode ultrasonic imaging algorithm. Finally, the simulation experiment of B-mode ultrasonic imaging was designed according to this algorithm, and the results showed that the sampling points required by the proposed FRI sampling model was 0.1% of traditional B-mode phased array ultrasonic imaging and the sampling frequency of the proposed ultrasonic imaging algorithm was 0.0077% of that of the traditional B-mode ultrasonic imaging scheme.

2. Basic Principle of Multichannel FRI Sampling

2.1. Multichannel FRI Direct Sampling Model. According to the FRI theory, for FRI signals, as long as K Fourier coefficients are known, the signals can be reconstructed according to the spectrum analysis. Therefore, getting the Fourier coefficient of FRI signals is the key of the problem.

The direct sampling model of Fourier coefficient is shown as Figure 1; T is the sampling period, that is, the time length of limited FRI signals; K is odd. For each channel, the original signal $x(t)$ is multiplied by a complex exponential function, which is the carrier signal; then its integral is divided by T to get the Fourier coefficient vector x of the channel. However, the model has an obvious disadvantage; that is, the frequency of the carrier function of each channel is different, which brings difficulties to circuit design.

2.2. Multichannel FRI Indirect Sampling Model. A Fourier coefficient indirect sampling model is shown as Figure 2, which is more practical than the direct sampling model. First, it has better robustness. The sampling result of each

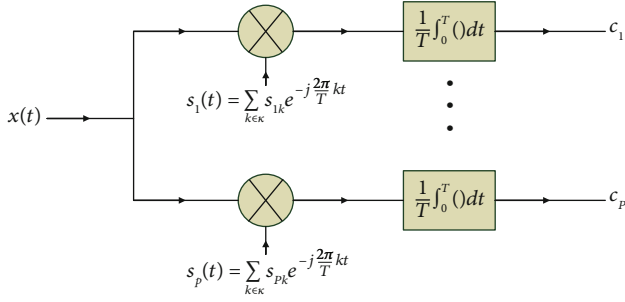


FIGURE 2: FRI indirect sampling model.

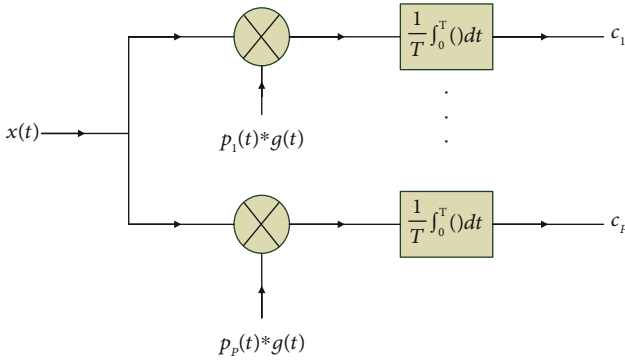


FIGURE 3: Multichannel FRI indirect sampling model of ultrasonic signal.

channel is the sum of a number of Fourier coefficients and will not affect the accuracy of the whole signal reconstruction due to the failure of one channel or more; second, it simplifies the carrier signal generation circuit, in which the carrier frequency components of each channel are the same, but the coefficients are different.

Its mathematical derivation is as follows:

Assume that there are P channels, and according to Figure 2, the carrier signal of channel i is

$$s_i(t) = \sum_{k \in K} s_{ik} e^{-j(2\pi/T)kt}, \quad (1)$$

where the coefficient $s_i(t) = \sum_{k \in K} s_{ik} e^{-j(2\pi/T)kt}$ of each channel is different; assume that the sampling vector $c = \{c_1, c_2, \dots, c_p\}$; then the sampling point of channel i is

$$c_i = \frac{1}{T} \int_0^T x(t) \sum_{k \in K} s_{ik} e^{-j(2\pi/T)kt} dt = \sum_{k \in K} s_{ik} x[k]. \quad (2)$$

Set S as the matrix of $P \times K$, the element (i, k) is S_{ik} , and then formula (2) can be written as the matrix form:

$$c = Sx, \quad (3)$$

where x is the Fourier coefficient vector and, if and only if column S is full rank, that is, $P \geq K$, S is left reversible; then,

$$x = S^+ c \quad (4)$$

Hybrid matrix S is determined by the selected carrier signal $S_i(t)$. Therefore, the Fourier coefficient direct sampling method is a special case of the Fourier coefficient indirect sampling method, where $P = K$ and $S = I$.

3. Design of Multichannel FRI Indirect Sampling Model of Ultrasonic Signal

Based on the practical application of FRI signal basic sampling model and ultrasonic signal, the designed sampling model is shown as Figure 3. In this model, the actual carrier signal $s_i(t)$ is generated from square wave signal $p_i(t)$ by means of filtering, and matrix A generated by square wave signal and hybrid matrix S generated by carrier signal are obtained by calculation.

3.1. Design of Square Signal $p_i(t)$. The mathematical model of square signal is shown as follows:

$$p_i(t) = \sum_{m \in \mathbb{Z}} \sum_{n=0}^{N-1} a_i[n] p\left(\frac{t-nT}{N-mT}\right), \quad (5)$$

where $a_i[n]$ is the coefficient vector whose length is N and its element can only be ± 1 ; set $P = N = K$; then $p(t)$ is

$$p(t) = \begin{cases} 1, & t \in \left[0, \frac{T}{N}\right], \\ 0, & t \notin \left[0, \frac{T}{N}\right]. \end{cases} \quad (6)$$

Its continuous Fourier transformation is

$$p(\omega) = \frac{T}{N} e^{-j(T/2N)\omega} \cdot \text{sinc}\left(\frac{T}{2\pi N}\omega\right). \quad (7)$$

3.2. Design of Carrier Signal $s_i(t)$. According to formula (5), $p_i(t)$ is a function with the period of T . Its Fourier efficient is shown as follows:

$$p_i(t) = \sum_{k \in \mathbb{Z}} d_i[k] e^{-j(2\pi/T)kt}, \quad (8)$$

where the Fourier coefficient $d_i[k]$ is worked out according to the following formula:

$$d_i[k] = \frac{1}{T} \int_0^T p_i(t) e^{-j(2\pi/T)kt} dt. \quad (9)$$

It can be seen that the formula is in the form of infinite addition; therefore, to get $s_i(t)$, a filter $g(t)$ must be added to filter $p_i(t)$:

$$s_i(t) = p_i(t) * g(t). \quad (10)$$

However, to achieve this function, $g(t)$ must have the following restrictions:

$$G(\omega) = \begin{cases} \text{nonzero,} & \omega = \frac{2\pi}{T}k, k \in \kappa, \\ 0, & \omega \neq \frac{2\pi}{T}k, k \notin \kappa, \\ \text{arbitrary,} & \text{elsewhere,} \end{cases} \quad (11)$$

where $G(\omega)$ is the continuous Fourier transformation of the filter $g(t)$ and $g(t)$ was selected as the ideal low-pass filter; we can get

$$\begin{aligned} s_i(t) &= \sum_{k \in \mathbb{Z}} d_i[k] G\left(\frac{2\pi}{T}k\right) e^{j(2\pi/T)kt} \\ &= \sum_{k \in \kappa} d_i[k] G\left(\frac{2\pi}{T}k\right) e^{j(2\pi/T)kt}. \end{aligned} \quad (12)$$

It can be seen from formulas (1) and (12) that

$$s_{ik} = d_i \left[-k - \left\lfloor \frac{K}{2} \right\rfloor \right] \cdot G\left(\frac{2\pi}{T} \left(- \left(k - \left\lfloor \frac{K}{2} \right\rfloor \right) \right) \right). \quad (13)$$

3.3. Design of Hybrid Matrix S. According to formulas (5) and (9), we can get

$$\begin{aligned} d_i[k] &= \frac{1}{T} \sum_{m \in \mathbb{Z}} \sum_{n=0}^N a_i[n] \int_0^T p\left(\frac{t-nT}{N-mT}\right) e^{-j(2\pi/T)kt} dt \\ &= \frac{1}{T} \sum_{n=0}^N a_i[n] \sum_{m \in \mathbb{Z}} \int_{-mT}^{-(m-1)T} p\left(\frac{t-nT}{N}\right) e^{-j(2\pi/T)kt} dt \\ &= \frac{1}{T} \sum_{n=0}^N a_i[n] \sum_{m \in \mathbb{Z}} \int_{-\infty}^{\infty} p\left(\frac{t-nT}{N}\right) e^{-j(2\pi/T)kt} dt \\ &= \frac{1}{T} \sum_{n=0}^N a_i[n] P\left(\frac{2\pi}{T}k\right) e^{-j(2\pi/T)kn}. \end{aligned} \quad (14)$$

With formulas (13) and (14), we can get

$$s_{ik} = \frac{1}{T} \sum_{n=0}^{N-1} a_i[n] P\left(\frac{2\pi}{T}k'\right) G\left(\frac{2\pi}{T}k'\right) e^{-j(2\pi/N)k'n}. \quad (15)$$

Then, formula (15) can be expressed with the hybrid matrix as

$$S = AW\Phi, \quad (16)$$

where A is the matrix of $P \times N$, the element $(i, n) A_{in} = a_i[n]$, and W is the diagonal matrix $w_{nk} e^{-j(2\pi/N)k'n}$ of the matrix of $N \times K$, where the diagonal element is

$$\Phi_{kk} = \frac{1}{T} P\left(\frac{2\pi}{T}k'\right) G\left(\frac{2\pi}{T}k'\right). \quad (17)$$

Substitute formula (7) into formula (17), we can get

$$\Phi_{kk} = \frac{1}{N} e^{-j(\pi/N)k'} \operatorname{sinc}\left(\frac{k'}{N}\right) G\left(\frac{2\pi}{T}k'\right). \quad (18)$$

When $g(t)$ is the ideal low-pass filter, formula (18) can be written as

$$\Phi_{kk} = \frac{1}{N} e^{-j(\pi/N)k'} \operatorname{sinc}\left(\frac{k'}{N}\right). \quad (19)$$

3.4. Design of Matrix A. Matrix A is generated according to the square signal. In the actual circuit, there is only one square signal generator, and the square signal on each channel is gotten by the delay of the square signal produced by the signal generator. In fact, the square signal on channel $(i+1)$ delays T/N (that is, a square wave) than the square signal i . As a rule, set $P = N = K = 2 \times L \times \text{oversampling} + 1$, where oversampling is the coefficient of oversampling. Then, matrix A can be designed as follows: the elements in the first row are written (can only be set as +1 or -1) from the square wave signal of the first channel. Start with the second row; the vector in the $(i+1)$ row is cyclically shifted from the i row vector to the left; for example, set $L = 3$, oversampling = 1, then $P = N = M = 7$; if the square wave in the first channel is shown as Figure 4, then matrix A can be written as formula (20), and the square signal and filtered carrier signal gotten by calculation are shown as Figure 5.

$$A_{7 \times 7} = \begin{pmatrix} 1 & -1 & 1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 & 1 & -1 & 1 \\ -1 & 1 & 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 \end{pmatrix}. \quad (20)$$

4. Mathematical Model of B-Mode Phased Array Ultrasonic Imaging

The B-mode ultrasonic imaging process is shown as Figure 6. Ultrasonic phased array probe is composed of several small wafers, which is called phased array element. During the transmission, the triggered signal is transmitted to the phased array controller, which will "command" the phased array unit to transmit an ultrasonic beam, which can be focused at different depths of detection. During the receiving process, the beam will be reflected back after encountering the obstacle; each phased array unit will process the received signal according to the calculated delay time and then adds up to get the reflected signal, which is

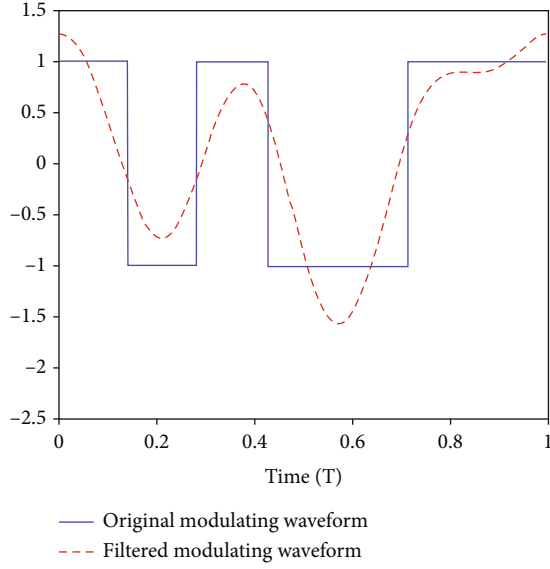


FIGURE 4: Square signal and filtered carrier signal.

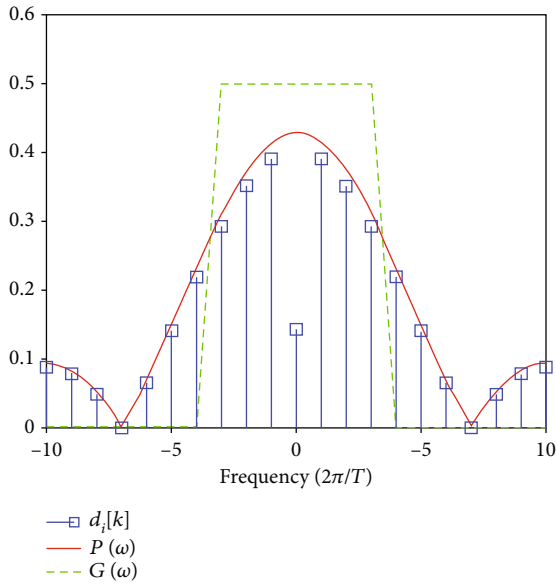


FIGURE 5: Square signal in the frequency domain and filtered carrier signal.

called a line scan. A two-dimensional image of the targeted area can be obtained for multiple line scans, which becomes B-mode ultrasonic imaging [39–41].

Figure 6 is the diagram of the line scan. $2M + 1$ phased array units distribute along the x -axis, the phased array unit m_0 was set as the origin, and δ_m expresses the distance from unit m to unit m_0 . The transmitted ultrasonic beam starts from the origin, and the included angle between the beam and z -axis is θ ; assume that the transmission time of the beam is 0, and when time $t > 0$, it arrives to the reflection point $(x, z) = (ct \sin \theta, ct \cos \theta)$, where c is the ultrasonic velocity. When the beam gets the reflection point, it will be received by other phased array unit except for returning

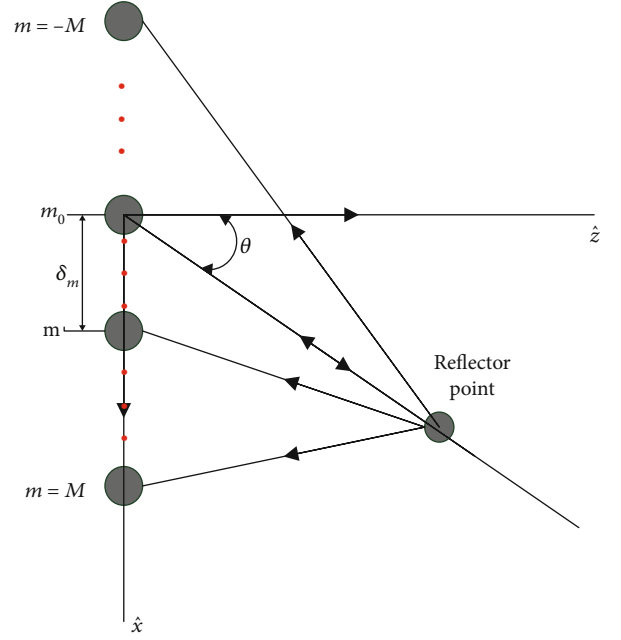


FIGURE 6: Line scan diagram.

units m_0 along the original path, where the distance of echo from the reflection point to the unit m should be

$$d_m(t; \theta) = \sqrt{(ct \cos \theta)^2 + (\delta_m - ct \sin \theta)^2}. \quad (21)$$

Therefore, it can be calculated that the time of echo from the reflection point to unit m is

$$\widehat{\tau}_m(t, \theta) = t + \frac{d_m(t; \theta)}{c} = t + \sqrt{t^2 + 4 \left(\frac{\delta_m}{c} \right)^2 - t \sin \theta}. \quad (22)$$

According to formula (22), the actual lag time of the signals received by each unit is different so as to align the signals on the time axis. Delay the signals $\varphi_m(t)$ received by each phased array unit for a period of time to obtain the adjusted signals:

$$\widehat{\varphi}_m(t; \theta) = \varphi_m \left(\frac{\widehat{\tau}_m(t; \theta)}{2} \right). \quad (23)$$

Then, we can get the following after the calculation:

$$\widehat{\varphi}_m(t; \theta) = \varphi_m \left(\frac{1}{2} \left(t + \sqrt{t^2 + 4 \left(\frac{\delta_m}{c} \right)^2 - t \sin \theta} \right) \right). \quad (24)$$

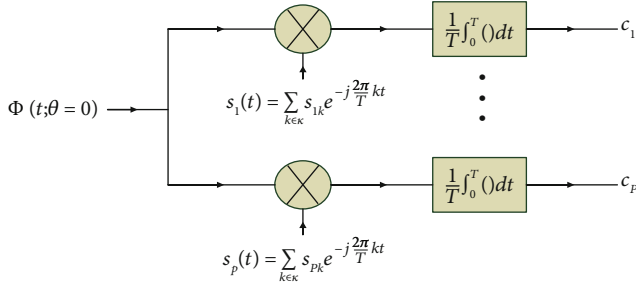


FIGURE 7: Ultrasonic signal multichannel FRI indirect sampling model.

When $\theta = 0$, the line scanning is called parallel scanning, and the scanning line is parallel to the axis. At this time, formula (24) can be written as

$$\widehat{\varphi}_m(t; \theta = 0) = \varphi_m \left(\frac{1}{2} \left(t + \sqrt{t^2 + 4 \left(\frac{\delta_m}{c} \right)^2} \right) \right). \quad (25)$$

At this time, δ_m represents the distance between unit m to the scanning line. Add all phased array unit $\widehat{\varphi}_m(t; \theta)$; then, we get the mathematical model of the line scanning:

$$\Phi(t; \theta) = \sum_{m=-M}^M \widehat{\varphi}_m(t; \theta) = \sum_{m=-M}^M \varphi_m \left(\frac{1}{2} \left(t + \sqrt{t^2 + 4 \left(\frac{\delta_m}{c} \right)^2} - t \sin \theta \right) \right), \quad (26)$$

where the mathematical model of the parallel scanning is

$$\Phi(t; \theta = 0) = \sum_{m=-M}^M \varphi_m \left(\frac{1}{2} \left(t + \sqrt{t^2 + 4 \left(\frac{\delta_m}{c} \right)^2} \right) \right). \quad (27)$$

5. Ultrasonic Imaging Algorithm Based on FRI Indirect Sampling

In the case of considering only parallel line scanning, since $\Phi(t; \theta = 0)$ is also a FRI signal, it is written as follows:

$$\Phi(t; \theta = 0) = \sum_{l=1}^L b_l h(t - t_l). \quad (28)$$

Indirect sampling of the Fourier coefficients is as shown in Figure 7.

We can get the sampling point of the channel q :

$$c_q = \frac{1}{T} \int_0^T \left\{ \sum_{k \in \kappa} \left(s_{qk} e^{-j(2\pi/T)kt} \right) \right\} \cdot \left\{ \sum_{m=-M}^M \varphi_m \left(\frac{1}{2} \left(t + \sqrt{t^2 + 4 \left(\frac{\delta_m}{c} \right)^2} \right) \right) \right\} dt. \quad (29)$$

After deformation, formula (29) can be written as

$$c_q = \sum_{m=-M}^M \left\{ \frac{1}{T} \int_0^{\widehat{T}} \widehat{s}_{q,m}(t) \varphi_m(t) dt \right\}, \quad (30)$$

where $\widehat{s}_{q,m}(t)$ is defined as follows:

$$\widehat{s}_{q,m}(t) \triangleq \left[1 + \left(\frac{\delta_m}{ct} \right)^2 \right] \left\{ \sum_{k \in \kappa} s_{qk} e^{-j(2\pi/T)k(t - (1/t)(\delta_m/c)^2)} \right\} \cdot u \left(t - \left| \frac{\delta_m}{c} \right| \right), \quad \text{for all } 1 \leq q \leq p, \quad -M \leq m \leq M, \quad (31)$$

where \widehat{T} is defined as follows:

$$\widehat{T} = \max_{m \in \{-M, \dots, M\}} \left\{ \frac{1}{2} \left(t + \sqrt{t^2 + 4 \left(\frac{\delta_m}{c} \right)^2} \right) \right\}, \quad (32)$$

where $u(t)$ is the step function paraded and it is defined as follows:

$$u(t) = \begin{cases} 1, & t \geq 0, \\ 0, & \text{else.} \end{cases} \quad (33)$$

Set

$$\widehat{t} = \frac{1}{2} \left(t + \sqrt{t^2 + 4 \left(\frac{\delta_m}{c} \right)^2} \right). \quad (34)$$

Then,

$$t = \widehat{t} - \frac{1}{\widehat{t}} \left(\frac{\delta_m}{c} \right)^2. \quad (35)$$

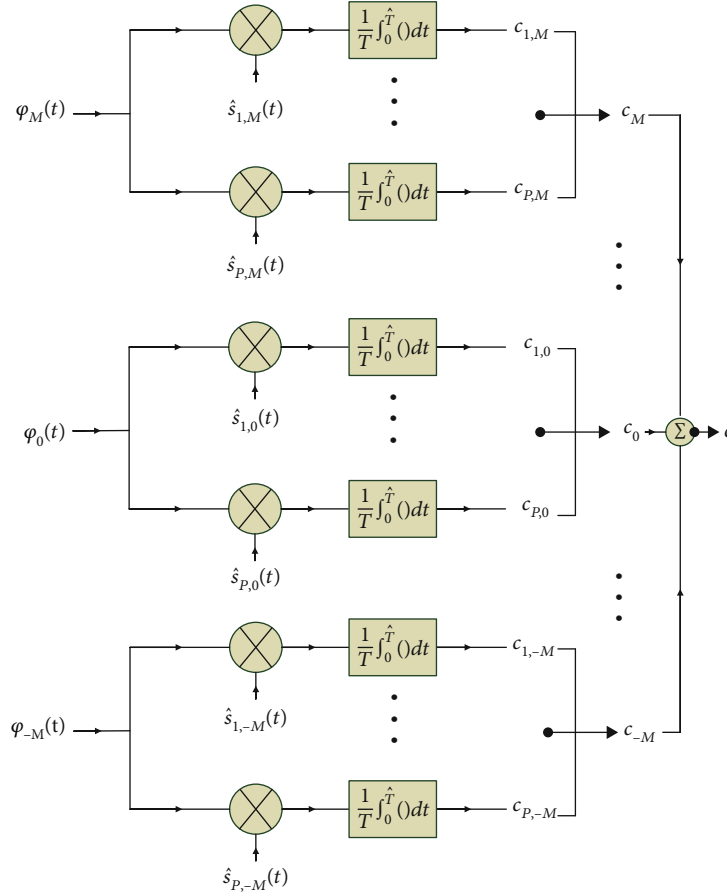


FIGURE 8: B-mode ultrasonic imaging FRI sampling model.

Substitute formula (35) into formula (29), then we can get

$$\begin{aligned}
 c_q &= \frac{1}{T} \int_0^{\hat{T}} \left\{ \sum_{k \in \kappa} \left(s_{q,k} e^{-j(2\pi/T)k(\hat{t} - (1/\hat{t})(\delta_m/c)^2)} \right) \right\} \\
 &\cdot \left\{ \sum_{m=-M}^M \varphi_m(\hat{t}) \right\} d\left(\hat{t} - \frac{1}{\hat{t}} \left(\frac{\delta_m}{c} \right)^2\right) \\
 &= \sum_{m=-M}^M \left\{ \frac{1}{T} \int_0^{\hat{T}} \left[1 + \left(\frac{\delta_m}{c\hat{t}} \right)^2 \right] \right\} \\
 &\cdot \left\{ \sum_{k \in \kappa} s_{q,k} e^{-j(2\pi/T)k(\hat{t} - (1/\hat{t})(\delta_m/c)^2)} \right\} \\
 &\cdot u\left(\hat{t} - \left| \frac{\delta_m}{c} \right|\right) \varphi_m(\hat{t}) d\hat{t}.
 \end{aligned} \quad (36)$$

Set $\hat{t} = t$; then write formula (36) as formula (30).

In conclusion, the mathematical model of B-mode ultrasonic imaging FRI sampling is shown as Figure 8.

The specific algorithm to realize the model is as follows:

- (1) Indirect sampling of the Fourier coefficients of the signals was received by each phased array unit, and

we can get $c_{q,m} = \int_0^{\hat{T}} \hat{s}_{q,m}(t) \varphi_m(t) dt$; define the sampling vector of length $Pasc_m = \{c_{1,m}, \dots, c_{p,m}\}$

- (2) After getting sampling vector of each phased array unit, add them; then, we get the final sampling vector $c = \sum_{m=-M}^M c_m$
- (3) Calculate Fourier coefficient vector $x = s^+ c$, where matrix s is $p \times k$ and its (i, k) element is s_{qk}
- (4) Use formula $y = H^{-1} x$; we can get the equation set $y_k = \sum_{l=1}^L a_l e^{-j2\pi k t_l t}$, $k \in \kappa$
- (5) Use Matrix Pencil to solve the equation set; we can get $\{t_l\}_{l=1}^L$, and the reason for using Matrix Pencil rather than Annihilating Filter to solve the equation set is that Matrix Pencil can estimate the number L of pulses
- (6) Substitute $\{t_l\}_{l=1}^L$ into the equation set; use the least square method to solve $\{b_l\}_{l=1}^L$
- (7) Use formula $\Phi(t; \theta = 0) = \sum_{l=1}^L b_l h(t - t_l)$ to reconstruct signal $\Phi(t; \theta = 0)$

6. Simulation

The emulation steps of the algorithm are as follows:

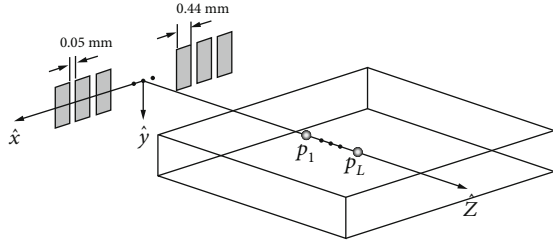


FIGURE 9: Working platform of ultrasonic imaging.

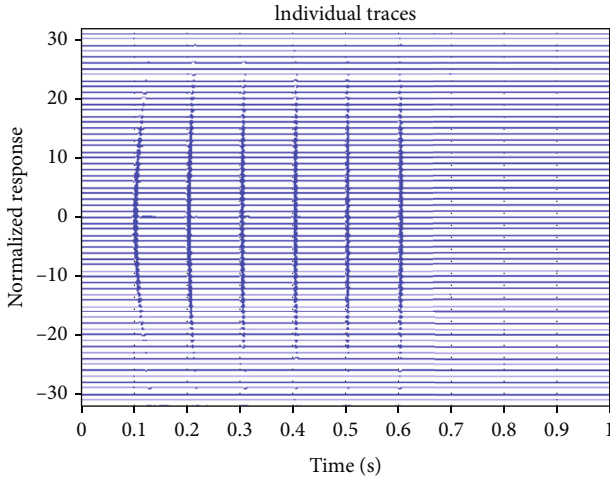


FIGURE 10: When $x = 0$, the original signals are on the 65 line arrays.

- (1) Use the software Field II to simulate the original signal of B-mode ultrasonic imaging, and the working platform is shown in Figure 9

The structure of the working platform is as follows. The platform consists of 65 line array units, marked as $\{m_{-32}, \dots, m_0, \dots, m_{32}\}$, the central frequency of the line array is $f_0 = 3.5$ MHz, the line array is arranged along axis \hat{x} , the width of the line array is $cf_0 = 0.44$ mm, and the height of it is 5 mm, measured by the axis \hat{y} ; the spacing of the line array is 0.05 mm, and the transmitting signal of the line array is the sine function of two periods, with a frequency of f_0 , that is $\sin(2\pi f_0 t)$, $0 \leq t \leq 2/f_0$; the focusing depth of the working platform is $r = 70$ mm, and the sampling frequency is $f_s = 100$ MHz.

- (2) The measured object is a cuboid $\{(x, y, z): |x| \leq 20 \text{ mm}, |y| \leq 5 \text{ mm}, 0 \leq z \leq 90 \text{ mm}\}$ including 6 reflection points evenly spaced along the axis \hat{z} : (0,0,10), (0,0,20), (0,0,30), (0,0,40), (0,0,50); and (0,0,60)
- (3) This emulation is parallel line scanning, that is $\theta = 0$, which is parallel to axis \hat{z} and is in plane $\hat{z}o\hat{x}$. Scan from $x = -20$ equally spaced to $x = 20$, with a total of 50 scans, and we get 50×65 original signals. What Figure 10 shows is the 65 original signals by line scanning (signals on the 65 line scanning unit) when $x = 0$

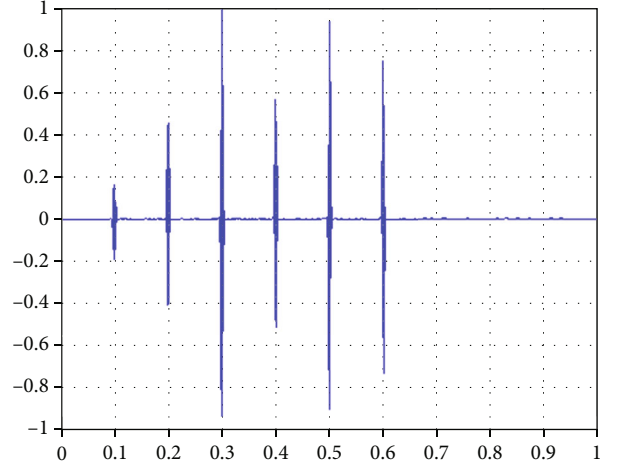


FIGURE 11: When $x = 0$, the original signals are on the line array unit.

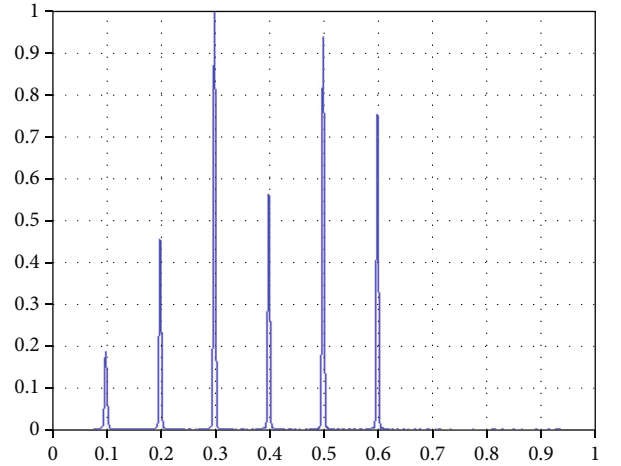


FIGURE 12: When $x = 0$, the modulated signal $\varphi_{m_0}(t)$ is on the line array unit.

- (4) When $x = 0$, the B-mode ultrasonic imaging process is as follows: as mentioned above, 65 signals are obtained, where the original signals of the line array units are shown as Figure 11

Since the original signal $h(t)$ is the sinusoidal pulse and its continuous Fourier transformation $H(w)$ are only two pulses. To ensure the reversibility of matrix H , the original signal needs to be modulated to make it become Gaussian pulse signal. The modulated signal $\varphi_{m_0}(t)$ is shown as Figure 12.

The modulated signal is gotten after modulating units on the 65 line array units; sampling vector c is obtained by using the sampling model shown in Figure 7, and $\{t_l\}_{l=1}^L$ is obtained after calculation. In ultrasonic imaging, getting the delay vector means getting the location information of the reflecting point. As shown in Figure 13, when the delayed vector $\Phi(t; \theta = 0)|_{x=0}$ is obtained, set the amplitude of all delayed vectors as 1.

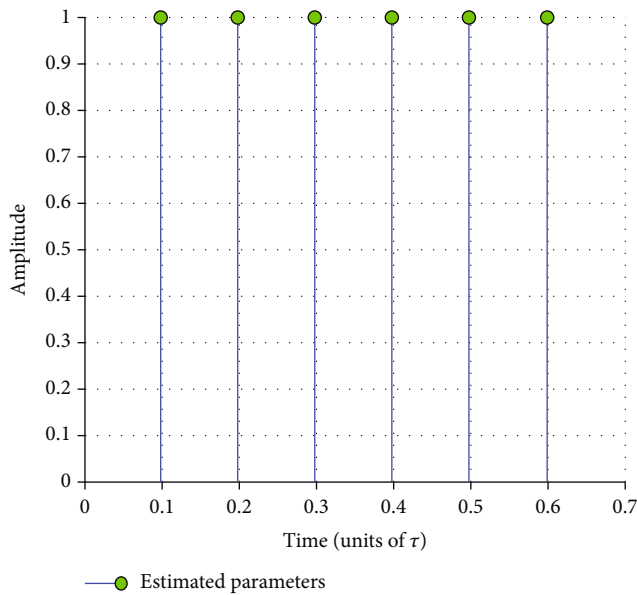


FIGURE 13: Delayed vector.

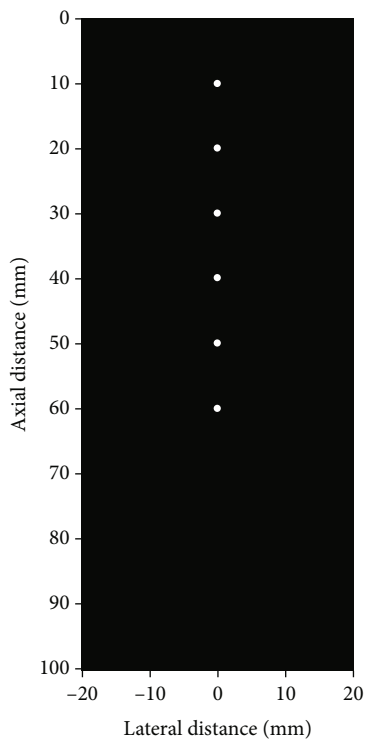


FIGURE 14: Reconstructed image.

- (5) Thus, repeat step (4) for 50 scanning lines; we get 50 groups of delayed vectors of $\Phi(t; \theta = 0)|_x$, $x = -20, -19.2, \dots, 0, \dots, 19.2, 20$
- (6) Use 50 groups of delayed vectors $\{t_l\}_{l=1}^6|_x$, and the reconstructed image is shown as Figure 14

According to the above analysis process, the number of sampling points required by traditional B-mode ultrasonic

imaging methods is $N_{\text{classic}} = (2R_{\text{max}}/c) \times f_s \times 50 = (2 \times 0.1/1540) \times (100 \times 10^6) \times 50 = 649350$, while the number of sampling points by using this method is $N_{\text{sub-Nyquist}} = (2L + 1) \times 50 = (2 \times 6 + 1) \times 50 = 650$.

Therefore, the sampling points based on Fourier coefficient indirect sampling model are $650/649350 = 0.1\%$ of that of traditional B-mode phased array ultrasonic imaging method, while the sampling frequency is $0.1/13 = 0.0077\%$.

7. Conclusion

- (1) A new FRI sampling model has been proposed based on ultrasonic signal. This model has the advantages such as good stability, simple circuit, and implementation
- (2) A B-mode phased array ultrasonic imaging algorithm has been proposed based on the FRI sampling model. This algorithm is proposed on the basis of both the FRI sampling model and the mathematical model characteristics of B-mode phased array ultrasonic imaging
- (3) The simulation program is designed for the algorithm, and the results show that the sampling point required by the proposed FRI sampling model is 0.1% of the traditional mode of B-mode phased array ultrasonic imaging, and the sampling frequency of the proposed ultrasonic imaging algorithm is 0.0077% of that of the traditional B-mode ultrasonic imaging

Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Natural Science Foundation of China under Grant 61671309, in part by the Guangdong Natural Science Foundation under Grant S2011010004487, and in part by the China Postdoctoral Science Foundation under Grant 2012M511551. The authors would like to thank Weiyi Lin, Chaoyi Dong, and Hongwei Sun for their contribution to this work.

References

- [1] S. Mallat and Guoshen Yu, "Super-resolution with sparse mixing estimators," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2889–2900, 2010.
- [2] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 1–8, Washington, DC USA, 2008.

- [3] Jianchao Yang, J. Wright, T. Huang, and Yi Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.
- [4] D. Friboulet, H. Liebgott, and R. Prost, "Compressive sensing for raw RF signals reconstruction in ultrasound," in *2010 IEEE International Ultrasonics Symposium*, pp. 367–370, San Diego, CA, USA, 2002.
- [5] C. Quinsac, A. Basarab, and D. Kouamé, "Frequency domain compressive sampling for ultrasound imaging," *Advances in Acoustics and Vibration*, vol. 2012, Article ID 231317, 16 pages, 2012.
- [6] H. Liebgott, R. Prost, and D. Friboulet, "Pre-beamformed RF signal reconstruction in medical ultrasound using compressive sensing," *Ultrasonics*, vol. 53, no. 2, pp. 525–533, 2013.
- [7] N. Dobigeon, A. Basarab, D. Kouamé, and J.-Y. Tourneret, "Regularized Bayesian compressed sensing in ultrasound imaging," in *Proc. European Signal Process Conf. (EUSIPCO)*, Bucharest, Romania, 2012.
- [8] C. Quinsac, F. de Vieilleville, A. Basarab, and D. Kouamé, "Compressed sensing of ultrasound single-orthant analytical signals," in *Proceedings of the IEEE International Ultrasonics Symposium (IUS'10)*, Orlando, FL, USA, 2010.
- [9] C. Quinsac, A. Basarab, D. Kouamé, and J.-M. Grégoire, "3D compressed sensing ultrasound imaging," in *Ultrasonics Symposium (IUS)*, pp. 363–366, San Diego, CA, USA, 2010.
- [10] C. Quinsac, A. Basarab, J.-M. Girault, and D. Kouamé, "Compressed sensing of ultrasound images: sampling of spatial and frequency domains," in *Signal Processing Systems (SIPS)*, pp. 231–236, San Francisco, CA, USA, 2010.
- [11] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, *Distributed Compressed Sensing*, Thermo Fisher, 2005.
- [12] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "An information-theoretic approach to distributed compressed sensing," in *Proc. 45rd Conference on Communication, Control, and Computing*, proc.allerton, 2005.
- [13] A. Basarab, H. Liebgott, O. Bernard, D. Friboulet, and D. Kouamé, "Medical ultrasound image reconstruction using distributed compressive sampling," in *Biomedical Imaging (ISBI)*, pp. 1945–7928, San Francisco, CA, USA, 2013.
- [14] H. Liebgott, A. Basarab, D. Kouame, O. Bernard, and D. Friboulet, "Compressive sensing in medical ultrasound," in *Ultrasonics symposium (IUS)*, pp. 1–6, Dresden, Germany, 2012.
- [15] A. Achim, B. Buxton, G. Tzagkarakis, and P. Tsakalides, "Compressive sensing for ultrasound RF echoes using a-stable distributions," in *Engineering in Medicine and Biology Society (EMBC)*, pp. 4304–4307, Buenos Aires, Argentina, 2010.
- [16] S. Hua, M. Yuchi, and M. Ding, "Compressed sensing for RF signal reconstruction in B-model ultrasound imaging," in *Intelligent Computation and Bio-Medical Instrumentation (ICBMI)*, pp. 19–22, Wuhan, China, 2011.
- [17] C. Quinsac, N. Dobigeon, A. Basarab, D. Kouamé, and J.-Y. Tourneret, "Bayesian compressed sensing in ultrasound imaging," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 101–104, San Juan, PR, USA, 2011.
- [18] N. Wagner, Y. C. Eldar, A. Feuer, G. Danin, and F. Z. Xampling, "Xampling in ultrasound imaging," in *SPIE Medical Imaging*, proc.allerton, 2011.
- [19] M. Mishali and Y. C. Eldar, "From theory to practice: sub-Nyquist sampling of sparse wideband analog signals," *IEEE Journal of Selected Topics on Signal Processing*, vol. 4, no. 2, pp. 375–391, 2010.
- [20] M. Mishali, Y. C. Eldar, and A. Elron, "Xampling: signal acquisition and processing in union of subspaces," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4719–4734, 2011.
- [21] T. Chernyakova and Y. C. Eldar, "Exploiting FRI signal structure for sub-Nyquist sampling and processing in medical ultrasound," in *Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, QLD, Australia, 2015.
- [22] T. Michaeli and Y. C. Eldar, "Xampling at the rate of innovation," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1121–1133, 2012.
- [23] R. Tur, Y. Eldar, and Z. Friedman, "Innovation rate sampling of pulse streams with application to ultrasound imaging," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1827–1842, 2011.
- [24] R. T. Gedalyahu and Y. C. Eldar, "Multichannel sampling of pulse streams at the rate of innovation," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1491–1504, 2011.
- [25] T. Chernyakova, Y. C. Eldar, and R. Amit, "Fourier domain beamforming for medical ultrasound," in *Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1520–6149, Vancouver, BC, Canada, 2013.
- [26] A. Burshtein, M. Birk, T. Chernyakova, A. Eilam, A. Kempinski, and Y. C. Eldar, "Sub-Nyquist sampling and Fourier domain beamforming in volumetric ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 5, pp. 703–716, 2016.
- [27] T. Chernyakova and Y. C. Eldar, "Fourier-domain beamforming: the path to compressed ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 61, no. 8, pp. 1252–1267, 2014.
- [28] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.
- [29] C. Bar-Zion, O. Tremblay-Darveau, D. A. Solomon, and Y. C. Eldar, "Super-resolution ultrasound imaging of vascular structures with high temporal resolution," *IEEE Transactions on Medical Imaging*, 2016.
- [30] O. Solomon, R. Cohen, Y. Zhang et al., "Deep unfolded robust PCA with application to clutter suppression in ultrasound," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 1051–1063, 2018.
- [31] R. Cohen and Y. C. Eldar, "Sparse convolutional beamforming for ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 12, pp. 2390–2406, 2018.
- [32] C. Bar-Zion, O. Tremblay-Darveau, D. A. Solomon, and Y. C. Eldar, "Fast vascular ultrasound imaging with enhanced spatial resolution and background rejection," *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp. 169–180, 2017.
- [33] M. Burshtein, T. Birk, A. Chernyakova, A. K. Eilam, and Y. C. Eldar, "Sub-Nyquist sampling and Fourier domain beamforming in volumetric ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 5, pp. 703–716, 2016.
- [34] M. Birk, A. Burshtein, T. Chernyakova et al., "Compressed 3D ultrasound imaging with 2D arrays," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6919–6923, Florence, Italy, May 2014.

- [35] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.
- [36] S. Guangming, C. Chongyu, J. Lin, X. Xie, and X. Chen, "Narrowband ultrasonic detection with high range resolution: separating echoes via compressed sensing and singular value decomposition," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, no. 10, pp. 2237–2253, 2012.
- [37] S. Guangming, C. Xuyang, S. Xiaoxia, Q. Fei, and A. Ding, "Signal matching wavelet for ultrasonic flaw detection in high background noise," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 58, pp. 776–787, 2011.
- [38] L. Yi, *Study on Ultrasonic Imaging Technology of Sparse Medicine*, Institute of Acoustics, Chinese Academy of Sciences, Beijing, 2013.
- [39] T. L. Szabo, *Diagnostics Ultrasound Imaging: Inside Out*, Elsevier Academic Press, Burlington, 2004.
- [40] J. A. Jensen, "Ultrasound imaging and its modeling," *Topics in Applied Physics*, vol. 84, no. 1, pp. 135–165, 2002.
- [41] J. A. Jensen, *Linear Description of Ultrasound Imaging Systems*, Technical University of Denmark: the International Summer School on Advanced Ultrasound Imaging, 1999.

Research Article

A Unified Calibration Paradigm for a Better Cuffless Blood Pressure Estimation with Modes of Elastic Tube and Vascular Elasticity

Jiang Shao,¹ Ping Shi ,¹ and Sijung Hu²

¹Institute of Rehabilitation Engineering and Technology, University of Shanghai for Science and Technology, Shanghai 200093, China

²Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Ashby Road, Loughborough, Leicestershire LE11 3TU, UK

Correspondence should be addressed to Ping Shi; garendon@163.com

Jiang Shao and Ping Shi contributed equally to this work.

Received 12 May 2020; Revised 27 February 2021; Accepted 17 March 2021; Published 13 April 2021

Academic Editor: Antonio Fernández-Caballero

Copyright © 2021 Jiang Shao 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.

Although two modes of elastic tube (ET) and vascular elasticity (VE) have been well explored for cuffless continuous blood pressure (BP) monitoring estimation, the initial calibration with these two models could be derived from different mathematical mechanisms for BP estimation. The study is aimed at evaluating the performance of VE and ET models by means of an advanced point-to-point (aPTP) pairing calibration. The cuff BPs were only taken up while the signals of PPG and ECG were synchronously acquired from individual subjects. Two popular VE models together with one representative ET model were designated to study aPTP as a unified assessment criterion. The VE model has demonstrated the stronger correlation r of 0.89 and 0.86 of SBP and DBP, respectively, and the lower estimated BP error of -0.01 ± 5.90 (4.55) mmHg and 0.04 ± 4.40 (3.38) mmHg of SBP and DBP, respectively, than the ET model. With the ET model, there is a significant difference between the methods of conventional least-square (LS) calibration and aPTP calibration ($p < 0.05$). These results showed that the VE model surpasses the ET model under the same uniform calibration. The outcome has been unveiled that the selection of initial calibration methods was vital to work out diastolic BP with the ET model. The study revealed an evident fact about initial sensitivity between the modes of different BP estimation and initial calibration.

1. Introduction

Uncontrolled hypertension or high blood pressure (BP) was a major risk factor that links to potential development of serious diseases such as stroke, hypertensive heart disease, and coronary artery disease [1]. For early warning, diagnosis, and treatment of hypertension in time, continuous cuffless BP monitoring technology was imperative. Conventional standard cuff-based BP measurement (for example, auscultation and oscillometry) was able to provide instantaneous information about BP status [2]. However, these cuff-based approaches with recurrently inflating and deflating of the cuff stress the patient, caused periodic interruptions to blood

flow, affected the physiological state of the patient, and disturb the quality of sleep due to repeat inflating and deflating of the cuff that stress the patient [3]. Consequently, the cuffless solution caught the attention of many scholars since it overcomes disturbance issues existing in the traditional cuff-based method [4, 5].

As a noninvasive optical measurement technique, photoplethysmography (PPG) provided valuable information on physiological heart monitoring and cardiovascular system assessment of vascular parameters [4, 6]. The pulse arrival time (PAT), defined as the time interval between the R peak and the point with maximum gradient on the rising edge of PPG, showed a high correlation with BP, especially systolic

blood pressure (SBP) [4, 6, 7]. In recent years, the PAT-based BP estimated models capturing BP variations mainly fall into two categories: vascular elasticity (VE) model [5, 6] and elastic tube (ET) model [8]. Among these reports [5, 6, 8], both the ET model and the VE model had excellent performance for BP estimation. The ET model, originated from the elastic Windkessel model, was continuously improved with the requirements of the elastic pipeline and gradually applied to the estimation and prediction of arterial BP [9].

Recently, the ET model based on PAT showed a better BP estimation performance than the VE model [8, 10]. However, the VE model based on the M-K equation was also widely studied and developed in cuffless continuous BP measurement due to it described the pulse wave velocity (PWV) in an infinitely thin-walled elastic vessel [5, 7]. Here, PWV was a measure of arterial stiffness which was a key predictor of future cardiovascular risk, or the rate at which pressure waves move down the vessel [11]. It has been established as a highly reliable prognostic parameter for cardiovascular morbidity and mortality in a variety of adult populations including older adults, patients with end-stage renal disease, diabetes, and hypertension [12]. For long-term cuffless BP monitoring technology, especially in home, a kind of accurate and practical model was particularly important. Importantly, the calibration method determined the practicality or the convenience of the specific BP estimation model. However, under the same calibration method, the performance evaluation of different models was ignored by the researchers. Therefore, it was necessary to use a unified calibration method to optimize BP models that were derived from different modeling methodologies, i.e., elastic tube (ET) theory and vascular elasticity (VE) theory to verify their effectiveness and applicability in estimation.

According to our previous work [13], the least-square (LS) calibration method (abbreviated as LS method or LS) [8] was usually applied in the ET model to accomplish the calibration procedure. This method is used to determine unknown parameters for a specific BP model in the short-time monitoring due to requiring all data sets regarding PAT obtained from ECG and PPG signals and cuff BPs. Consequently, consecutive long-term monitoring could not be accurately completed and achieved. To our knowledge [13], the initial BP monitoring period should not be ignored in early warning, diagnosis, and treatment of unhealthy physical conditions. More importantly, the accuracy of BP estimation depends on the sample/point size for this LS method. Obviously, it was a great limitation to complete continuous BP monitoring using LS methods to obtain unknown coefficients or parameters in a specific BP estimation model for each subject.

By comparing with the LS method, one sample point-to-point (oPTP) pairing method (abbreviated as the oPTP method or oPTP) [5, 6] is a calibration technique that uses a one-to-one mapping relationship between model function and variable/sample to determine unknown parameters of the specific model, usually effectively adopted for calibration in the VE model, only needed one point to complete the whole calibration step. The oPTP method to some extent overcame the limitation of larger data samples in the LS method. However, the oPTP method demanded highly for one point/sample in the initial personalized calibration

procedure. It was important to highlight the fact that the selection of the initial point determined the quality of the BP estimation. The VE model required in a quiet state, however, the parameters, like PAT and BP, always fluctuated in a range of small variations at rest, as influenced by the accuracy of the VE model for BP estimation.

In this study, an advanced point-to-point (aPTP) pairing calibration method (abbreviated as the aPTP method or aPTP) was proposed to examine and access the effectiveness, the accuracy, and the robustness of BP estimation compared with the cumbersome LS method and sensitive oPTP method. The investigation in this study includes the following: (1) the correlation and overall performances between the cuff BP and the estimated BP were examined in a uniform calibration method, i.e., aPTP; (2) aPTP was studied to verify whether it could replace the LS method in personalized procedure; (3) for the ET model, the LS method and aPTP method were both studied to compare their effectiveness and the applicability for BP estimation; (4) for the VE model, BP estimation based on the aPTP method was investigated to verify whether possessing high accuracy and robustness compared with sensitive oPTP.

The purpose of the study is to evaluate the performance of VE and ET models by means of an aPTP pairing calibration method. With the study, a cost-effective cuffless BP monitoring approach could be emerged with an easy and durable personalized calibration. Such an approach could be anticipated to be a better choice when considering the practicality of long-term and continuous BP monitoring with both modes of elastic tube and vascular elasticity. Besides these, the study has proofed an evidence about the sensitivity of BP estimation along with these models and their initial calibration methods.

2. Materials and Methods

2.1. Methods

2.1.1. Modeling Methodology for BP Estimation. The beating heart created BP and flow pulsations that propagate as waves through the arterial tree, and then, the waves were reflected at transitions in arterial geometry and elasticity [14]. As a hemodynamic parameter, arterial BP fluctuated on a beat-to-beat basis due to the dynamic interplay from vasomotion, neural regulation, and arterial mechanisms [6]. It was physiologically affected by four factors: arterial compliance, cardiac output, peripheral resistance, and blood volume [15]. Arterial compliance was evaluated by PAT since it was an index of arterial stiffness [16]. In regard to peripheral resistance and blood volume, one of the primary sources was the change in arterial diameter [17]. In recent years, the PAT-based BP estimated models mainly included two categories: elastic tube (ET) model and vascular elasticity (VE) model.

The ET model was developed from the theory of elastic tubes. It was based on two important assumptions and premises: (i) the blood vessel was equivalent to an elastic tube, and (ii) the compliance of the arterial system remained constant throughout the cardiac cycle [18]. Introducing the blood pulsation information and giving the arterial

compliance (C) into the transmission line of the pressure wave, the ET model was developed and proposed. Here, a novel BP estimation nonlinear model was derived by Esmaili et al. from the conservation of mass and momentum principle equation, called the M-M model [8], a representative ET mode. Arterial compliance as an important quantity with respect to these assumptions, used in physiology, was the degree to which a container experiences pressure or force without disruption [14]. It depended strongly on pressure. Considering the conservation of mass and momentum equations, it could be seen that C was a function of pressure on walls of blood vessels, i.e., BP.

The VE model was linked and established according to the Moens-Korteweg (M-K) equation [4]. In biomechanics, the M-K equation modeled a relationship between the wave speed or pulse wave velocity (PWV) and the incremental elastic modulus (a coefficient of elasticity) of the arterial wall or its distensibility [19]. It involved two assumptions: (i) the artery wall was thin and was modeled as a thin shell, and (ii) the thickness and radius of the artery still fixed as the blood pressure changes [20]. Additionally, PWV as an important quantity in the M-K equation was commonly used as a clinical marker of vascular elasticity [12]. Combining the M-K equation with an exponential arterial elasticity model [21, 22], the MK-EE model as a new BP-PAT model was obtained. It gave a logarithmic relationship between BP and the PAT. Moreover, a new BP-PAT model, i.e., MK-BH model [2], was introduced from the Bramwell-Hill (B-H) equation [5] to consider the nonlinear nature in the cardiovascular system. Based on the MK-BH model, Zheng et al. established a mathematical relationship between MBP and a factor that reflected the change in elasticity caused by pressure wave variations, which was called the dMK-BH model [5]. The modeling principle and mechanism about the above three representative models are expressed in Table 1.

2.1.2. A Unified Calibration Paradigm: Advanced Point-to-Point (aPTP) Calibration. As mentioned previously, the conventional LS method in the ET model required all data sets in the whole process of BP monitoring to accomplish the personalized procedure in the corresponding ET model that was derived from the theory of elastic tubes. Consequently, this could be extremely troublesome to implement this procedure. Remarkably, the popular oPTP calibration method only required one point/sample to obtain unknown parameters in the specific VE model that was derived from the theory of vascular elasticity. It is thus clear that one lone sample used to complete the initial calibration process is sensitive as such BP calibration may be accidental and inaccurate.

Here, a new aPTP method was proposed to overcome oPTP method's initial sensitivity and access BP estimation property under a unified paradigm for models that were derived from different modeling methodologies. The mapping relationship between dependent variable and independent variable was established through the available initial values. This technique was called the point-to-point pairing calibration method, i.e., one cuff BP value pairing with PAT_{mean} , a parameter with the average value of PAT. The advanced PTP (aPTP) calibration method (shown in Figure 1) was

developed from the traditional PTP calibration method in the present study.

Three steps of (1) the initial calibration processing, (2) the robust control strategy, and (3) the average treatment effect were established up as the aPTP method.

Step 1. Initial calibration processing.

Generally, the digital cuff-type BP monitor will obtain a set of SBP and DBP after each inflation and deflation. During this period, a series of PAT samples can be calculated according to the ECG and PPG signals detected by the sensors, that is, $PAT_1, PAT_2, \dots, PAT_l$. Here, l is the number of heartbeats during the inflation and deflation of the cuff sphygmomanometer. In this way, we can calculate the average value of this series of PAT samples as follows:

$$PAT_{\text{mean}} = \frac{1}{l} \sum_{i=1}^l PAT_i. \quad (1)$$

If there are N undetermined parameters ($N = 1$ and 3 for the VE and ET models in this study, respectively) among the BP estimation model, then N cuff BP (including SBP and DBP) and N mean PAT (i.e., PAT_{mean}) need to be paired one-to-one to obtain the values of the undetermined parameters. Here, it is defined as θ , that is, $SBP_0, DBP_0, PAT_0, a_i,$ and b_i . This pairing relationship can be understood as solving the inverse function of θ from the relationship $\mathbf{BP} = f(\theta; \mathbf{PAT})$ between BP and PAT, as follows:

$$\theta = f^{-1}(\mathbf{BP}, \mathbf{PAT}) \quad (2)$$

where θ , \mathbf{BP} , and \mathbf{PAT} are m -row and n -column matrices. The f denotes the one-to-one mapping relationship between BP and PAT. In addition, m and n denote the number of subjects and the undetermined parameters θ , respectively, in the BP estimation model. After this step, one determined parameter θ will be obtained.

Step 2. The robust control strategy.

Given the possibility that cuff BP values of subjects in a quiet state might be the same, in the present study, a robust control strategy is necessary to guarantee the validity and rigor of calibration for obtaining all parameter values in a specific BP model. In this regard, we propose two robust control strategies: function analytical solution definition and numerical floating control. The former is to determine whether each obtained model parameter θ_i is a real number, which ensures that this model parameter θ_i is valid in step 1. The latter is to determine whether each θ_i is different. This strategy will traverse whole θ_i obtained from step 1 in the resting state, which to some extent expands the limitation of the conventional oPTP pairing calibration method in the sample or point and the sensitivity of BP estimation. After this step, a set of determined parameters θ_i will be obtained.

TABLE 1: Summary of mathematical models to calculate BP from PAT.

Models	SBP	DBP	Category
M-M [8]	$a_1 + \sqrt{b_1 + c_1} * \frac{1}{PAT^2}$	$a'_1 + \sqrt{b'_1 + c'_1} * \frac{1}{PAT^2}$	ET model
MK-BH [2]	$SBP_0 - \frac{2}{\gamma * PAT_0} (PAT - PAT_0)$	$SBP - (SBP_0 - DBP_0) * \left(\frac{PAT_0}{PAT}\right)^2$	VE model
dMK-BH [5]	$DBP + (SBP_0 - DBP_0) * \left(\frac{PAT_0}{PAT}\right)^2$	$\frac{SBP_0}{3} + \frac{2DBP_0}{3} + \frac{2}{\gamma} \ln\left(\frac{PAT_0}{PAT}\right) - \frac{SBP_0 - DBP_0}{3} * \left(\frac{PAT_0}{PAT}\right)^2$	

Note 1: γ denoted a vascular information parameter. Note 2: SBP_0 , DBP_0 , and PAT_0 were the base value of SBP, DBP, and PAT, respectively, and could be determined at the beginning of monitoring by calibration using an additional cuff-type BP monitor device. Note 3: a_i , b_i , a'_i , b'_i ($i = 1, 2$); c_1 , c'_1 were the corresponding function coefficients.

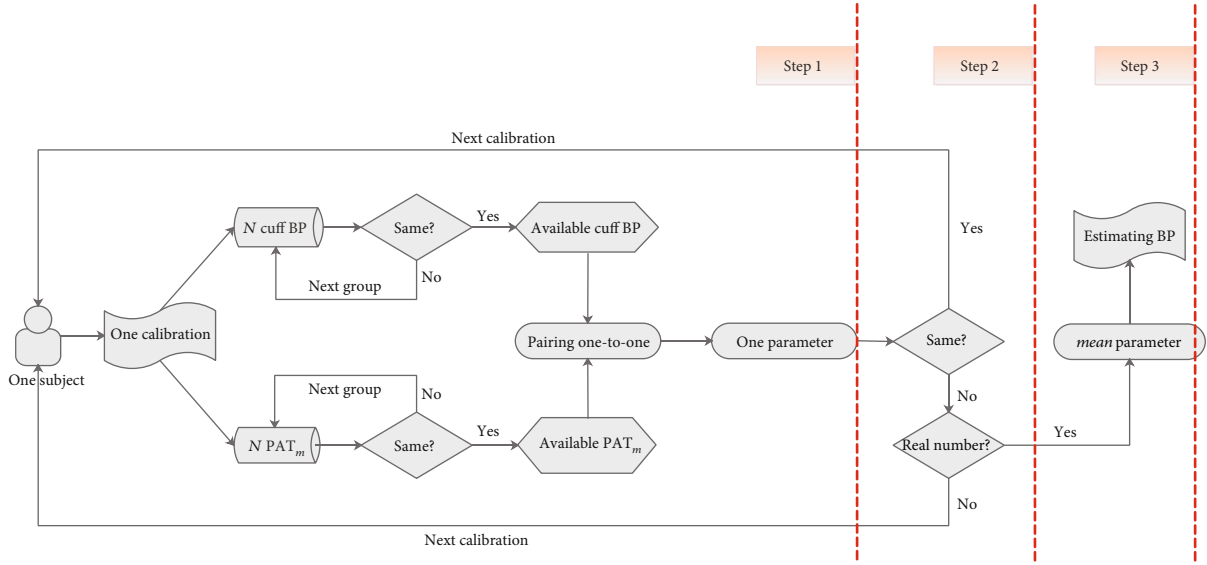


FIGURE 1: The aPTP calculation procedure for the BP monitoring system.

Step 3. The average treatment effect.

Finally, the average values of these parameters of their respective BP estimation models were taken as the final BP monitoring parameters, i.e., SBP_0 , DBP_0 , PAT_0 , a_i , and b_i , as follows:

$$\hat{\theta} = \frac{1}{M} \sum_{i=1}^M \theta_i, \quad (3)$$

where M denotes the number of available model parameter θ_i . The data collection of cuff BP and PAT from BP monitor and sensors, respectively, in the whole calibration process took about eight minutes, and the subjects were required to keep peace and quiet. This calibration process was done only one time per subject, and after deriving parameters in the BP estimator model, the BP could be estimated continuously. That is, using mentioned calibrated parameters $\hat{\theta}$ and their respective nonlinear models introduced in Subsection 2.1.1, SBP and DBP are estimated.

2.1.3. A Summary of Different Calibration Methods and BP Models in Terms of Mechanism.

The LS method and oPTP method were employed for the initial calibration of the ET model and VE model, respectively. The aPTP method could be directly applied to different BP models. And there was no need for extra requirements about distinguishing the modeling mechanism of the models for aPTP. More importantly, it only employed limited data sets at rest rather than all data sets throughout the process of BP monitoring to complete the personalized calibration procedure. The respective application of different initial calibration methods in two BP estimation modes is listed in Table 2. To go a step further, the relationship between the three calibration methods and respective performance is given in Figure 2 to elucidate them at a clear level.

2.2. Experimental Protocol.

This experimental protocol was performed in a study room with temperature $22.6 \pm 2.3^\circ\text{C}$ and relative humidity 60–70%. The PowerLab/16sp system (Castle Hill, ADInstruments, Australia, 2002) was employed to synchronously record and amplify the ECG and PPG signals. The ECG signal was filtered by a 1 Hz high-pass filter and a 40 Hz low-pass filter. Meanwhile, the PPG signal was filtered by a 0.5 Hz high-pass filter and a 20 Hz low-pass filter, and the sampling frequency was 1 kHz [8]. Since the PPG sensor was placed on the subject's left hand, the cuff-type

TABLE 2: Summary of initial calibration methods and mathematical models.

Measurements	oPTP	aPTP	LS	Category and mechanism	
M-M [8]	○	✓	✓	Nonlinear	Elastic tube (ET) model
MK-BH [2]	✓	✓	○	Linear, nonlinear	Vascular elasticity (VE)
dMK-BH [5]	✓	✓	○	Nonlinear	

Note 1: “✓”: applicable; “○”: not applicable. Note 2: a linear model for SBP estimation in the MK-BH model.

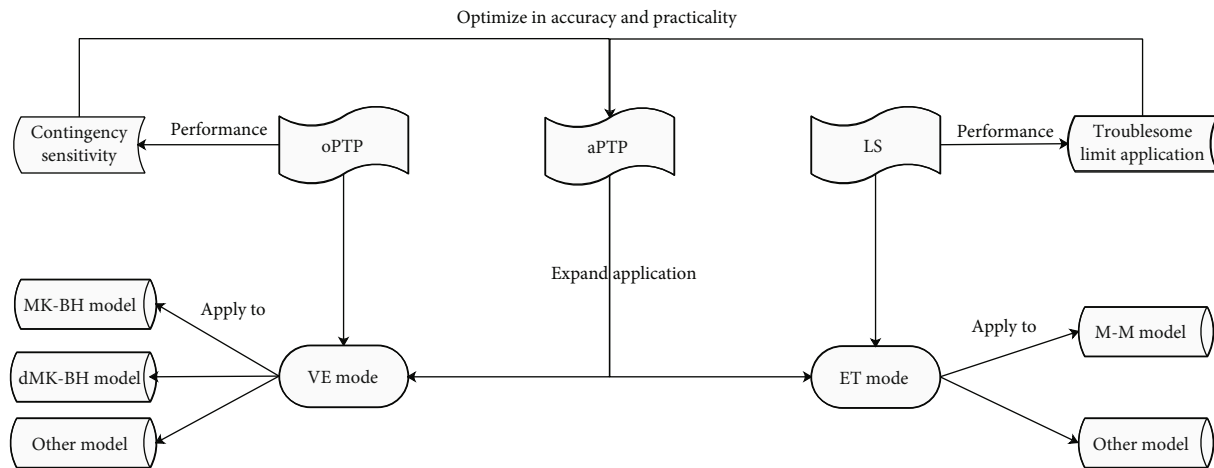


FIGURE 2: The relationship between the three calibration methods and respective performance.

BP monitor (MB-300C, Jasun, China) was mounted on the right arms of a subject to provide a reference BP reading and reduce the effects of BP measurement on the PPG signal. Referring to the guidelines of the cuff-type BP monitor, the accuracy of the cuff-type BP monitor was ± 3 mmHg to be in compliance with the clinical golden standards of AAMI [23]. Specifically, the mean absolute error of less than 5 mmHg and the standard deviation (SD) of mean error of less than 8 mmHg (i.e., the difference should be within 5 ± 8 mmHg) were considered as an acceptable error rate referred to the AAMI guidelines [23].

It should be pointed that before starting the data collection process, we measured the BP of six subjects randomly using a cuff-type BP monitor (MB-300C, Jasun, China) and a conventional mercury sphygmomanometer with a rigorous experimental process. Obviously, the measured BPs were approximately the same for each subject (each person was at the rest or peace condition during BP measurement, so approximately the same values for BPs for each person were expected) [8]. Here, using these two devices, the mean absolute errors (MAEs) of SBP and DBP measurement values were 2.7 and 3.2 mmHg for six subjects, respectively.

2.3. Data Acquisition Procedure and Data Analysis. Twelve healthy subjects in the age range of 21-37 years (9 males and 3 females) without a history of cardiovascular or neurological disorders participated in this study. All participants gave written informed consent. The study was approved by the health center authorities at the University of Shanghai for Science and Technology.

Among these common BP interventions [4, 24, 25], a designated physical exercise is employed since it has been shown

to cause a sensible increase in both SBP and DBP up to 40 mmHg [7]. Currently, the same supervised physical exercise, which was climbing 12 floors at a constant rate for five minutes, was used for all the subjects to guarantee a greater change in BP to obtain a more accurate model estimation [26]. Just finishing the physical exercise, the cuff BP, the ECG, and the PPG signals were collected. Each subject with the cuff BP measurement was asked to sit upright on a chair 25 cm away from the table and breathe naturally to avoid the motion artifact interference. A total of data collection took around 15 minutes per subject after physical exercise.

Generally, the BP estimation based on PAT from a period of 30 s cuff BP, ECG, and PPG signals does not begin until the initial calibration procedure has been completed. In this study, a total of 365 pairs of cuff BP vs. PAT_{mean} , a parameter with the average value of PAT, data sets from at least 30,000 heartbeats were tested for twelve subjects. The estimated errors between the cuff BP and the estimated BP were evaluated as the mean error (mean) \pm standard deviations (SD) as well as the mean absolute difference (MAD). They were defined as follows:

$$\begin{aligned}
 \text{mean} &= \frac{1}{n} \sum_{i=1}^n (BP_{\text{est}_i} - BP_{\text{cuff}_i}), \\
 \text{SD} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (BP_{\text{est}_i} - BP_{\text{cuff}_i})^2}, \\
 \text{MAD} &= \frac{1}{n} \sum_{i=1}^n |BP_{\text{est}_i} - BP_{\text{cuff}_i}|, \\
 \text{SSE} &= \sum_{i=1}^n (BP_{\text{est}_i} - BP_{\text{cuff}_i})^2, \\
 \text{RMSE} &= \sqrt{\frac{\text{SSE}}{n}}.
 \end{aligned} \tag{4}$$

where BP_{est_i} and BP_{cuff_i} denoted the i th BP measured through BP estimation models and by the reference cuff method, respectively, and n was the number of measured BP used for evaluation.

3. Results

Under a unified calibration paradigm, i.e., aPTP method, the ET model and VE model based on PAT were investigated from correlation and overall performance to assess their property for BP estimation. Further, we performed difference analysis to test whether the aPTP method showed better convenience than the cumbersome LS method for the ET model and higher robustness than the sensitive oPTP method for the VE model. More specially, we compared the estimated values of the model derived from each subject data with the corresponding reference values for the model derived from cuff BP for each subject through statistical analysis and regression analysis to ascertain whether there was any difference among the aPTP method and LS method for the ET model. Similarly, we also investigated the property of BP estimation under the conventional oPTP method and new proposed aPTP method for the VE model.

3.1. Assess the Performance under a Unified Paradigm for Different Modeling Methodology

3.1.1. Correlation. To compare the estimated BP results from the three most popular functions (M-M, MK-BH, and dMK-BH) quantitatively, we computed Pearson's correlation coefficient (r), the summed square of residuals (SSE), and root mean squared error (RMSE) between the cuff BP and the estimated BP for all subjects.

As shown in Table 3, a relatively high correlation with SBP was observed among the ET and VE models. The M-M model, as a nonlinear ET model, had the largest SSE and RMSE between both SBP and DBP compared to others. Moreover, the M-M model showed the weakest correlation coefficient with DBP ($r = 0.74$) compared with others. For the VE model, MK-BH and dMK-BH had higher correlation coefficients with BP than the ET model, i.e., M-M model. Remarkably, the dMK-BH had the highest correlation coefficient with BP ($r = 0.89$ for SBP and $r = 0.86$ for DBP) compared to others. More significantly, the dMK-BH model developed from the MK-BH model showed the smallest SEE and RMSE for estimated BP compared with the M-M model.

3.1.2. Overall Performances. The criteria for overall performance evaluation included mean errors of estimation, MAD of estimation, and SD of estimation. Moreover, the average value and 95% confidence intervals of BP estimated error were calculated to identify the influence of different models on the accuracy of estimated BP. A Kruskal-Wallis test with Dunn's multiple comparison test was executed to determine whether statistically significant differences were observed between the mean errors of the ET model and the VE model. These performances are shown in Figure 3.

According to Figure 3, the BP could not be properly estimated from the ET model compared to the VE model. This

M-M model had a mean \pm SD (MAD) of 1.11 ± 7.51 (5.57) mmHg for SBP and -0.23 ± 6.47 (5.13) mmHg for DBP estimated error, respectively, while the dMK-BH as a nonlinear model had a mean \pm SD (MAD) of -0.01 ± 5.90 (4.55) mmHg for SBP and 0.04 ± 4.40 (3.38) mmHg for DBP, respectively. For the VE model, the MAD of estimated errors in the dMK-BH model was decreased by 0.93 and 1.31 mmHg compared with MK-BH. Remarkably, the SD of the errors for all methods was within 8 mmHg for SBP and DBP. It was consistent with the AAMI requirements of 5 ± 8 mmHg/7 mmHg (mean \pm SD/MAD) for the BP estimated error rate [23]. Additionally, for SBP estimation using the aPTP method, a significant difference between the ET model and the VE model was not observed. In contrast with SBP estimation, there was a significant difference between the M-M model and the MK-BH model for DBP estimation ($p < 0.01$). Similarly, a significant difference between the MK-BH model and the dMK-BH model was also observed for DBP estimation ($p < 0.0001$).

3.2. Difference's Analysis Using Different Calibration Methods among the ET and VE Models

3.2.1. BP Estimation Using the Cumbersome LS Method and Convenient aPTP Method for the ET Model. Here, differences were tested with Kruskal-Wallis tests and with Dunn's multiple comparison tests to determine whether statistically significant differences between cuff BP and estimated BP obtained from the ET model using the LS method and aPTP method. More details are plotted in Figure 4.

According to Figure 4, there was a significant difference between the traditional LS method and the unified aPTP method for BP estimation using the ET model ($p < 0.05$). However, it was noteworthy that there were no significant changes on cuff BP and estimated BP based on the ET model. Particularly, there were no difference changes between the cuff BP and estimated BP based on the ET model by using the LS calibration method. Similarly, the difference changes were not found between the cuff BP and estimated BP based on the ET model by using the aPTP calibration method.

Moreover, for the ET model, we also investigated the correlation between the cuff BP and the estimated BP obtained by using LS and PTP calibration methods to test and verify whether using the LS method might reinforce the BP estimation performance. More details about regression plots are shown in Figure 5.

In Figure 5, based on the LS method that required all samples for BP estimation, r values obtained from all the subjects were 0.86 and 0.83 for SBP and DBP estimations, respectively. Regarding the aPTP method that only required small samples, r values for SBP and DBP estimations were 0.83 and 0.74, respectively. Consequently, under different calibration methods, the ET model showed larger differences in the performance of BP estimation, especially in DBP estimation.

3.2.2. BP Estimation Using the Sensitive oPTP Method and Robust aPTP Method for the VE Model. For the VE model, BP estimation using the unified aPTP method was investigated

TABLE 3: BP results using the BP models of ET and VE.

Models	SBP			DBP			Category
	r	SSE	RMSE	r	SSE	RMSE	
M-M [8]	0.83	14940	6.57	0.74	10690	5.56	Elastic tube (ET) model
MK-BH [2]	0.81	11610	5.79	0.77	8672	5.01	Vascular elasticity (VE) model
dMK-BH [5]	0.89	10180	5.42	0.86	5370	3.94	

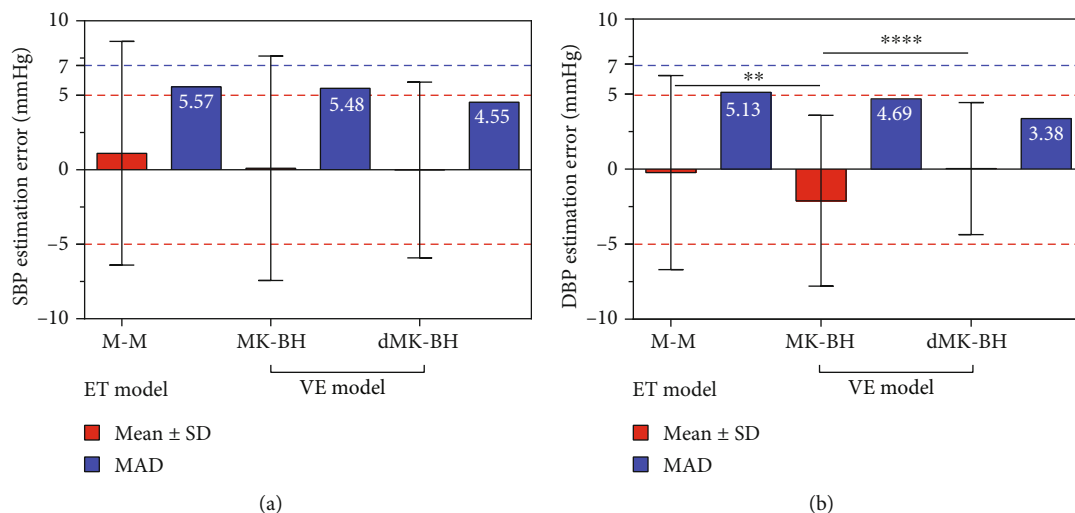


FIGURE 3: The overall comparison of different methods for (a) SBP and (b) DBP measurement. Note 1: the red and blue dotted lines denoted the largest boundary for mean error (5 mmHg) and MAD (7 mmHg). Note 2: significant differences: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

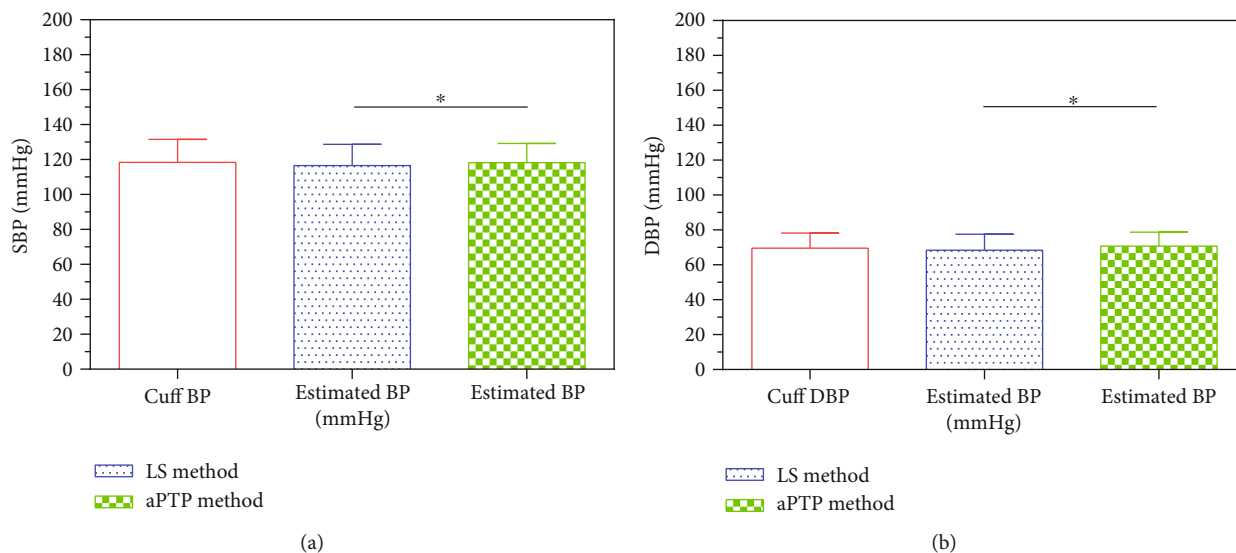


FIGURE 4: Difference comparisons between cuff BP and estimated BP based on the ET model by using LS and aPTP calibration methods. Significant differences: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

to verify whether possessing high accuracy and robustness compared with sensitive oPTP. Here, we computed the r , SSE, and RMSE between cuff BP and estimated BP. Details of each estimated BP function are reported in Table 4 for the MK-BH model and Table 5 for the dMK-BH model.

According to Table 4, for the MK-BH model, using the unified aPTP method that serves as an initial calibration method to estimate BP showed better performance than the traditional oPTP method. Compared with the oPTP method, a stronger correlation coefficient with SBP ($r = 0.81$) and SBP

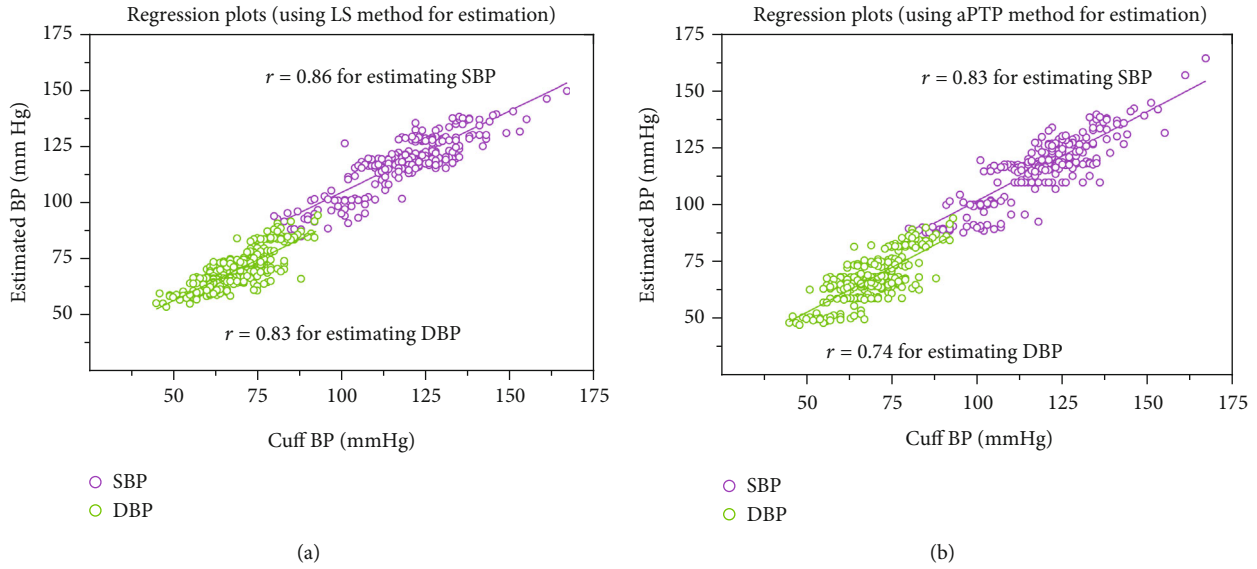


FIGURE 5: The correlation comparison of BP estimation based on the ET model by using the (a) LS method and (b) aPTP method.

($r = 0.77$) and the larger values of SEE and RMSE were observed. Similar performance also was found in the dMK-BH model according to Table 5.

Regarding AAMI [23], a BP device must comply with the constraints of mean error < 5 mmHg, the SD of the error < 8 mmHg, and the MAD of the error < 7 mmHg between the new technique and the reference. Hence, it was necessary to investigate the estimated errors of the VE model using the oPTP and aPTP methods to test whether the estimated BP based on a new proposed calibration method was acceptable under AAMI criteria. Here, differences were tested with Kruskal-Wallis tests and with Dunn's multiple comparison tests in Figure 6.

As observed from Figure 6, the estimated errors of the VE model, i.e., MK-BH and dMK-BH model, meet the AAMI requirements. More importantly, using the aPTP method for BP estimation showed high accuracy comparing with the oPTP method. For instance, the estimated error values of SD decreased by approximately 3 mmHg for the MK-BH model and dMK-BH model, respectively. For the MK-BH model, the estimated errors of MAD cut by 1.60 mmHg and 0.85 mmHg for SBP and DBP estimation when using the unified aPTP method. Similarly, for the dMK-BH model, the estimated errors between cuff BP and estimated BP had almost the same decline as the MK-BH model in the case of using the aPTP method. For DBP estimation, a statistically significant difference between the oPTP method and the aPTP method was observed for MK-BH ($p < 0.0001$) and dMK-BH ($p < 0.01$), respectively, and there was also a statistically significant difference between the MK-BH and dMK-BH models for oPTP ($p < 0.0001$) and aPTP ($p < 0.0001$), respectively. Note that difference changes were not found in SBP estimation.

3.3. Comparison with Prior Works. Our study achieved comparable results to the rest of the studies. Here, all results are presented in both mean \pm SD (MAD) and Pearson's correlation coefficient (i.e., r). Table 6 presents a comparison of the

results reported in this paper with the results reported in the literature.

Some comparisons regarding initial calibration methods could be made according to Table 6. Compared with the previous researches, the experimental samples in the present investigation were appropriately selected. Specifically, the experimental samples of 365 pairs of cuff BP vs. PAT_{mean} from at least 30,000 heartbeats in the present work were larger than investigations of Chen et al. [27], Esmaili et al. [8], Proença et al. [28], and Tang et al. [2], with the experimental samples of 200, 173, 166, and 169 pairs of cuff BP vs. PAT_{mean} , respectively. Also, the age range of 12 healthy subjects from 22 to 37 years old was appropriate according to Tang et al.'s work [2] and Huynh et al.'s investigation [29], with the age range of 12 healthy subjects from 20 to 31 years old and 15 healthy subjects from 24 to 34 years old, respectively.

In addition, Table 6 also provides more details on calibration methods, signal collection methods, and BP estimation errors. For instance, Huynh et al. [29] proposed a revised PTP calibration method by using three pairs of BP for BP estimation and achieved an accuracy of 0.31 ± 8.55 (6) mmHg and -0.5 ± 5.07 (5) mmHg for the estimated SBP and DBP, respectively. Similarly, Zheng et al. [30] selected oPTP as an initial calibration procedure and utilized ECG and PPG signals to estimate BP, reporting the error of 2.4 ± 5.7 (6) mmHg for the SBP estimation, and no errors regarding DBP were investigated. Baek et al. [31] further proposed a multiple regression of PAT, HR, and TDB (a kind of arterial stiffness index defined as the duration from the maximum derivative point to the dirotic peak in the PPG signal) [32] for BP estimation and achieved an accuracy of -0.02 ± 7.04 (5.50) and 0.00 ± 5.08 (3.86) mmHg for the estimated SBP and DBP, respectively. Recently, Simjanoska et al. [33] developed a probability distribution method to accomplish personalized procedure and reported the SD (MAD) of 10.22 (7.72) and 10.03 (9.45) for the estimated SBP and DBP, respectively.

TABLE 4: The performance of the MK-BH model using oPTP and aPTP calibration methods.

Methods	SBP			DBP		
	r	SSE (mmHg)	RMSE (mmHg)	r	SSE (mmHg)	RMSE (mmHg)
oPTP	0.72	16220	6.848	0.73	10930	5.619
aPTP	0.81	11610	5.792	0.77	8672	5.006

TABLE 5: The performance of the dMK-BH model using oPTP and aPTP calibration methods.

Methods	SBP			DBP		
	r	SSE (mmHg)	RMSE (mmHg)	r	SSE (mmHg)	RMSE (mmHg)
oPTP	0.87	15790	6.755	0.75	8747	5.028
aPTP	0.89	10180	5.4.25	0.86	5386	3.945

4. Discussion

In this study, a called aPTP method was proposed to unify the paradigm of personalized procedure for BP estimation models (i.e., ET and VE models, see Subsection 2.1.1) deriving from different methodology mechanisms. Comparing with the cumbersome LS method and sensitive oPTP method, the effectiveness, the accuracy, and the robustness of BP estimation were further investigated to validate the property of using the aPTP method in the different ET model (M-M) and VE model (MK-BH and dMK-BH). According to our investigation, different personalized calibration methods showed large differences in both ET and VE models, respectively. Not only that, the performance of BP estimation was also quite different in the same BP model, i.e., ET or VE model. These findings were particularly evident in DBP estimation. Moreover, there is no significant difference for SBP estimation (see Figures 3, 4, and 6). By contrast, a significant difference for DBP estimation was observed (see Figures 3, 4, and 6) in both the BP model (M-M, MK-BH, and dMK-BH) and calibration methods (LS, oPTP, and aPTP). For instance, the correlation coefficients and the values of SSE and RMSE were different among ET and VE models using LS and aPTP methods (see Table 3), especially in estimating DBP for the ET model. These results warned us that more attention should be paid to the selection of initial calibration methods when estimating DBP for the ET model.

For the ET model, using the previous LS method significantly enhanced the correlation in the case of DBP compared with using the aPTP method (see Figure 5). A strong correlation in DBP was of great importance, since generally, in the literature [2, 5, 8, 21], correlation coefficients of DBP estimations were distinctly less than those of SBP estimations. Consequently, the LS method might be an effective method to achieve a strong correlation in DBP. At this point, we presented an evidence on using LS and aPTP methods would lead to different BP estimation performance for the ET model based on PAT obtained from ECG and PPG signals. However, the LS method could not meet the requirement of a small initial sample size, for example, some samples obtained

from 5-minute signals [21, 34] in personalized calibration procedure because it acquires all data sets for long-term BP monitoring. Although the aPTP method was slightly weaker than the LS method in the correlation between cuff BP and estimated BP, using aPTP to finish personalized calibration procedure for BP estimation still meets the AAMI criteria for the ET model and VE models, respectively (see Figure 3) due to better performance of BP estimation (see Figures 3 and 5; Tables 3, 4, and 5). Hence, using the same and uniform calibration method, for example, aPTP method, was confirmed to be necessary when comparing the property of BP estimation under BP models deriving from different modeling methodologies.

Key information to be observed is that the ET model (i.e., M-M) had larger estimated errors between cuff BP and estimated BP than the VE model under selecting the aPTP method as the initial calibration method (see Figures 3 and 5; Tables 3, 4, and 5). As mentioned previously, the ET model was developed from the theory of elastic tubes supposing the blood vessel was equivalent to an elastic tube [8]. In fact, the actual arterial system contained branches, which were elastically and geometrically taper and terminated within the microcirculation, rather than a simple tube [4]. Therefore, research on the influence of the vascular branches on the M-M model was interesting and necessary in the next study.

Regarding the VE model, i.e., MK-BH and dMK-BH, the performance of BP estimation has been investigated to assess the accuracy, the effectiveness, and the robustness using the conventional oPTP method and the proposed aPTP method. For the aPTP method, MK-BH and dMK-BH models showed greater property in terms of BP estimation than the oPTP method (see Subsection 3.2.2). The select aPTP using small samples (for example, data set obtained from 5-minute signals) [21, 34] is recommended as the initial unified calibration method by comparison with the LS method using all data sets as the consequence of convenience in personalized calibration procedure (see Figure 1). The outcome gained from the aPTP method has been expressed to be applied in the ET model and showed a good accuracy and effectiveness of BP estimation with meeting the AAMI requirement [23] (see Figures 3 and 4; Table 3).

Here, it was necessary to point out that no matter which calibration method was used, the dMK-BH model was superior to MK-BH in BP estimation (see Tables 3, 4, and 5; Figures 3 and 6). The basis of the modeling sources indicated the variability of BP estimation performance. As mentioned previously, the MK-BH model [2] was proposed based on the B-H equation and M-K equation to strengthen the correlation between the estimated DBP and cuff DBP. Furthermore, the dMK-BH model was developed from the MK-BH model through introducing MBP for better estimate BP. Hence, the introduction of MBP to the dMK-BH model might be a main reason of its more accuracy for BP estimation than the MK-BH model. This reveals that MBP was a key factor in the cuffless BP estimation model. However, as reported in some literature [5, 21, 27], the sensitive coefficient γ in the dMK-BH model limited its practicality to a great extent. Moreover, the vascular information parameter γ changed with aging and the development of cardiovascular

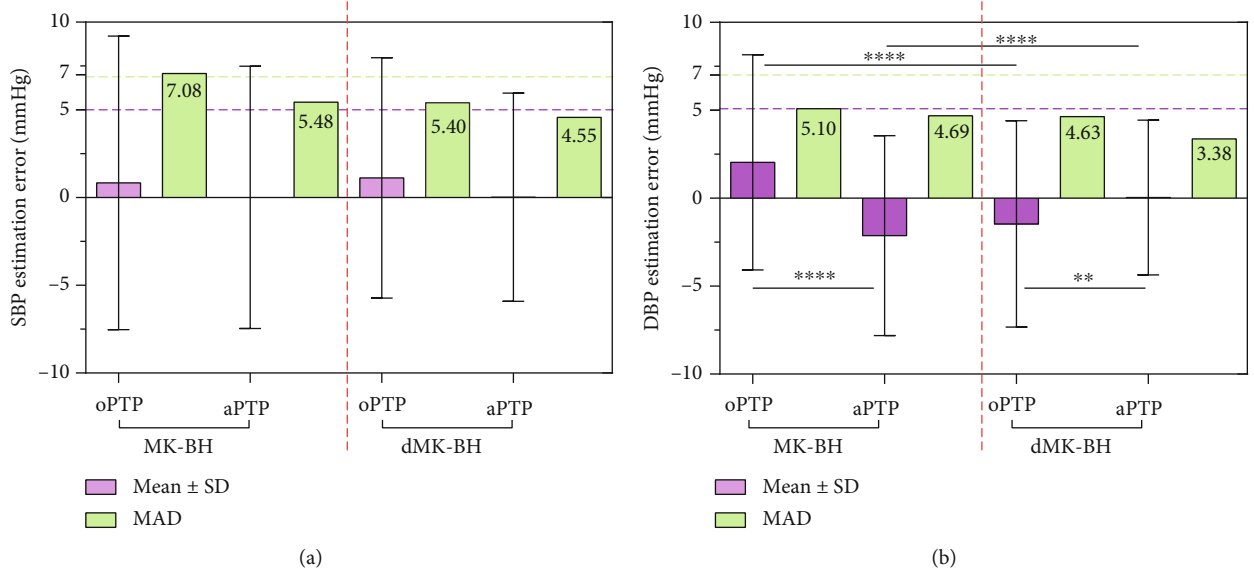


FIGURE 6: The overall performance comparison using oPTP and aPTP methods based on VE models to estimate (a) SBP and (b) DBP, respectively. Note: the purple and green dotted lines denoted the largest boundary for mean error (5 mmHg) and MAD (7 mmHg). Significant differences: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

TABLE 6: Comparison results with prior work.

Calibration method	Acquired signals (measure location)	Subjects	Test samples (pairs of BP)	Accuracy w.r.t. cuff BP (mean \pm SD (MAD)/mmHg; r)
LS [27]	Only PPG: ear & toe	$N = 20$	200	SBP: 0 ± 8 (7) [†] ; / DBP: 0 ± 8 (7) [†] ; /
LS [8]	ECG: hands & leg; PPG: finger; PCG: chest	$N = 32$	173	SBP: 0.12 ± 6.15 (4.71); 0.95 DBP: 1.31 ± 5.36 (4.44); 0.84
LS [28]	ECG: thorax; EIT: thorax	$N = 24$	166	SPAP: 0.7 ± 3.8 [‡] (6 [†]); 0.87 [‡]
PTP, three pairs [29]	PPG: finger; IPG: wrist	$N = 15$	90	SBP: 0.31 ± 8.55 (6 [†]); 0.88 DBP: -0.5 ± 5.07 (5 [†]); 0.88
oPTP [2]	ECG: palms; PPG: finger	$N = 12$	169	SBP: 0.2 ± 5.8 (4.4); 0.89 DBP: 0.4 ± 5.7 (4.6); 0.83
oPTP [30]	ECG: arm; PPG: arm	$N = 10$	70	SBP: 2.4 ± 5.7 (6 [†]); 0.80 DBP: no estimation
Multiple regression [31]	ECG: back; PPG: thigh; BCG: thigh	$N = 5$	1147	SBP: -0.02 ± 7.04 (5.50); 0.86 DBP: 0.00 ± 5.08 (3.86); 0.81
Probability distributions [32]	Only ECG: chest	$N = 51$	3219	SBP: $/\pm 10.22$ (7.72); / DBP: $/\pm 10.03$ (9.45); /
aPTP, this work (MK-BH model)	ECG: wrist & foot PPG: finger	$N = 12$	365	SBP: -0.77 ± 7.79 (6.0); 0.83 DBP: 0.51 ± 5.70 (4.4); 0.74
aPTP, this work (dMK-BH model)	ECG: wrist & foot PPG: finger	$N = 12$	365	SBP: -0.54 ± 6.95 (5.3); 0.87 DBP: 0.24 ± 5.21 (4.0); 0.78

Note 1: “/” = not be estimated based on reported results or able to be reported in corresponding authors’ other work. Note 2: “†” = be approximately estimated from the corresponding Bland-Altman plots. Note 3: “‡” = the median value of the corresponding estimation accuracy index. Note 4: PCG = phonocardiogram; BCG = ballistocardiography; IPG = impedance plethysmography; EIT = electrical impedance tomography; SPAP = systolic pulmonary artery pressure.

diseases [22]. Therefore, it was not easy to obtain an optimal γ value in different ages and pathophysiologic conditions. Remarkably, pulse transmit time and photoplethysmogram intensity ratio (PIR) [6, 17] were recently suggested to apply in the establishment of the dMK-BH model for BP estimation to achieve better performance of BP estimation than before.

Referring to the preliminary outcome from this study, we are confident that the aPTP method as an effective calibration method could be used for ambulatory and home BP monitoring to some extent in the future. Furthermore, more in-depth measurements including ECG [4, 32], PPG [4, 6], BCG [31], IPG [29], PCG [8], and others [32] need to be involved in

future research to fully verify this aPTP calibration method. Considering that physiological functions such as vascular elasticity and vascular sizes of different individuals will change with time, periodic calibration should be considered to improve the reliability of BP measurement through recalibrating parameters in a specific BP estimation model. Recently, introducing some covariates including HR [34, 35], PWV [36], and PIR [6, 17] and extra relevant variables [37] into personalized calibration procedures was expected for achieving a better prediction of BPs.

Two limitations were also found to estimate BP. One limitation was that the practical application of the M-K equation using the VE model implied several assumptions [5, 6, 8, 14]. For example, one assumption was that the thickness to radius ratio [20, 36] was a constant, which led to the invalidity for complex behaviors and regulation of the involved arterial tree. Moreover, arterial segments involved in BP estimation were formed for both elastic and muscular arteries with different biomechanical properties. The influence of these factors on BP estimation was left for the further study. Similarly, for the ET model, the simple tube needs to be greatly improved to make this specific tube have more performance of actual arterial system contained branches for each subject. Another limitation was that the subjects were generally young and healthy. Hence, further studies with extensive validation that included a larger population of individuals recruited from different age groups were required to confirm and extend these conclusions. Meanwhile, some novel models including the description of the prejection period (PEP) were worth to be established due to the VE and ET models which were not well considered with the influence of PEP and vascular tone changes.

5. Conclusions

A called advanced point-to-point (aPTP) pairing calibration method was proposed to unify the paradigm of personalized procedure for two modes (VE and ET) of BP estimation models deriving from different methodology mechanisms. Comparing with the cumbersome LS method and the sensitive oPTP method, the outline of aPTP pairing calibration is with the following:

- (1) Characteristics: the aPTP method requiring small samples or points can improve the robustness and accuracy of the initial calibration technique in BP monitoring. At the same level, like using the aPTP method, the arterial VE model based on the M-K equation was superior to the ET model developed from the nonlinear theory of elastic tubes
- (2) Applicable scope: the aPTP method was made available for both VE and ET models. Its three-step calibration strategy provided a calibration paradigm for almost all BP estimation models. What is more, evidence was provided about sensitivity in both calibration methods and BP models
- (3) Further work: more in-depth measurements including ECG, PPG, ballistocardiography, and impedance

plethysmography are required to be involved in future research to fully verify and enhance this work

All in all, the aPTP method was expected to unify the initial calibration method under different BP models and achieve an easy and durable personalized calibration procedure for cost-effective cuffless BP monitoring technology.

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 no conflict of interest.

Authors' Contributions

Jiang Shao and Ping Shi contributed equally to this work.

Acknowledgments

The authors would like to thank all the participants in this study.

References

- [1] S. Lewington, R. Clarke, and N. Qizilbash, "Age-specific relevance of usual BP to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies," *Lancet*, vol. 361, no. 9366, pp. 1389–1390, 2002.
- [2] Z. Tang, T. Tamura, M. Sekine et al., "A chair-based unobtrusive cuffless blood pressure monitoring system based on pulse arrival time," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 5, pp. 194–205, 2016.
- [3] E. Kaniusas, H. Pftzner, L. Mehnen et al., "Method for continuous nondisturbing monitoring of blood pressure by magnetoelastic skin curvature sensor and ECG," *IEEE Sensors Journal*, vol. 6, no. 3, pp. 819–828, 2006.
- [4] R. Muckamala, J. O. Hahn, O. T. Inan et al., "Toward ubiquitous blood pressure monitoring via pulse transit time: theory and practice," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 8, pp. 1879–1901, 2015.
- [5] Y. Zheng, C. C. Y. Poon, B. P. Yan, and J. Y. W. Lau, "Pulse arrival time based cuff-less and 24-H wearable BP monitoring and its diagnostic value in hypertension," *Journal of Medical Systems*, vol. 40, no. 9, pp. 1–11, 2016.
- [6] X. Ding, B. P. Yan, Y.-T. Zhang, J. Liu, N. Zhao, and H. K. Tsang, "Pulse transit time based continuous cuffless blood pressure estimation: a new extension and a comprehensive evaluation," *Scientific Reports*, vol. 7, no. 1, article 11554, 2017.
- [7] Y. M. Wong, C. Y. Poon, and Y. T. Zhang, "An evaluation of the cuffless blood pressure estimation based on pulse transit time technique: a half year study on normotensive subjects," *Cardiovascular Engineering*, vol. 9, no. 1, pp. 32–38, 2009.
- [8] A. Esmaili, M. Kachuee, and M. Shabany, "Nonlinear cuffless blood pressure estimation of healthy subjects using pulse transit time and arrival time," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 12, pp. 3299–3308, 2017.

- [9] N. Westerhof, J. W. Lankhaar, and B. E. Westerhof, "The arterial windkessel," *Medical & Biological Engineering & Computing*, vol. 47, no. 2, pp. 131–141, 2009.
- [10] D. L. Newman and S. E. Greenwald, *Validity of the Moens-Korteweg equation*, Springer Berlin Heidelberg, 1978.
- [11] J. Bank and D. R. Kaiser, "Smooth muscle relaxation: effects on arterial compliance, distensibility, elastic modulus, and pulse wave velocity," *Hypertension*, vol. 32, no. 2, pp. 356–359, 1998.
- [12] R. R. Townsend, I. B. Wilkinson, E. L. Schiffrin et al., "Recommendations for improving and standardizing vascular research on arterial stiffness," *Hypertension*, vol. 66, no. 3, pp. 698–722, 2015.
- [13] J. Shao, P. Shi, S. Hu, and H. Yu, "A revised point-to-point calibration approach with adaptive errors correction to weaken initial sensitivity of cuff-less blood pressure estimation," *Sensors*, vol. 20, no. 8, pp. 1–15, 2020.
- [14] F. N. Van de Vosse and N. Stergiopoulos, "Pulse wave propagation in the arterial tree," *Annual Review of Fluid Mechanics*, vol. 43, no. 1, pp. 467–499, 2011.
- [15] "Factors that affect blood pressure," http://www.edises.it/file/minicd/germ002/misc/assignmentfiles/cardiovascular/Fact_Aff_Blood_Pressure.pdf.
- [16] P. Birutė, S. Daukantas, and V. Marozas, "Assessment of pulse arrival time for arterial stiffness monitoring on body composition scales," *Computers in Biology and Medicine*, vol. 85, pp. 135–142, 2016.
- [17] X. R. Ding, Y. T. Zhang, J. Liu, W. X. Dai, and H. K. Tsang, "Continuous cuffless blood pressure estimation using pulse transit time and photoplethysmogram intensity ratio," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 5, pp. 964–972, 2016.
- [18] K. S. Heffernan, S. Y. Jae, K. R. Wilund, J. A. Woods, and B. Fernhall, "Racial differences in central blood pressure and vascular function in young men," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 295, no. 6, pp. H2380–H2387, 2008.
- [19] D. J. Korteweg, "Ueber die Fortpflanzungsgeschwindigkeit des Schalles in elastischen Röhren," *Annual Review of Physical Chemistry*, vol. 241, no. 12, pp. 525–542, 2006.
- [20] Y. Ma, J. Choi, A. Hourlier-Fargette et al., "Relation between blood pressure and pulse wave velocity for human arteries," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 44, pp. 11144–11149, 2018.
- [21] W. Chen, T. Kobayashi, S. Ichikawa, Y. Takeuchi, and T. Togawa, "Continuous estimation of systolic blood pressure using the pulse arrival time and intermittent calibration," *Medical & Biological Engineering & Computing*, vol. 38, no. 5, pp. 569–574, 2000.
- [22] D. J. Hughes, C. F. Babbs, L. A. Geddes, and J. D. Bourland, "Measurements of Young's modulus of elasticity of the canine aorta with ultrasound," *Ultrasonic Imaging*, vol. 1, no. 4, pp. 356–367, 1979.
- [23] Association for the advancement of medical instrumentation [AAMI], *American national standard: non-invasive sphygmomanometers - part 2: clinical investigation of intermittent automated measurement type*, ANSI/AAMI/ISO 81060-2:2019, 2021, <http://webstore.ansi.org/Standards/AAMI/ANSIAAMIISO810602019>.
- [24] C. N. Joseph, C. Porta, G. Casucci et al., "Slow breathing improves arterial baroreflex sensitivity and decreases blood pressure in essential hypertension," *Hypertension*, vol. 46, no. 4, pp. 714–718, 2005.
- [25] J. Y. A. Parati, C. S. Lim, and P. Wang, "Evaluation of blood pressure changes using vascular transit time," *Physiological Measurement*, vol. 27, no. 8, pp. 685–694, 2006.
- [26] Noordergraaf, *Circulatory System Dynamics, Vol. 1*, Elsevier, Amsterdam, The Netherlands, 2012.
- [27] Y. Chen, C. Wen, G. Tao, M. Bi, and G. Li, "Continuous and noninvasive blood pressure measurement: a novel modeling methodology of the relationship between blood pressure and pulse wave velocity," *Annals of Biomedical Engineering*, vol. 37, no. 11, pp. 2222–2233, 2009.
- [28] M. Proença, F. Braun, M. Lemay et al., "Non-invasive pulmonary artery pressure estimation by electrical impedance tomography in a controlled hypoxemia study in healthy subjects," *Scientific Reports*, vol. 10, no. 1, article 21462, 2020.
- [29] T. H. Huynh, R. Jafari, and W. Y. Chung, "Noninvasive cuffless blood pressure estimation using pulse transit time and impedance plethysmography," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 4, pp. 967–976, 2019.
- [30] Y.-L. Zheng, B. P. Yan, Y.-T. Zhang, and C. C. Y. Poon, "An armband wearable device for overnight and cuff-less blood pressure measurement," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 7, pp. 2179–2186, 2014.
- [31] H. J. Baek, G. S. Chung, K. K. Kim, and K. S. Park, "A smart health monitoring chair for noninvasive measurement of biological signals," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 1, pp. 150–158, 2012.
- [32] H. J. Baek, K. K. Kim, J. S. Kim, B. Lee, and K. S. Park, "Enhancing the estimation of blood pressure using pulse arrival time and two confounding factors," *Physiological Measurement*, vol. 31, no. 2, pp. 145–157, 2010.
- [33] M. Simjanoska, M. Gjoreski, M. Gams, and A. M. Bogdanova, "Non-invasive blood pressure estimation from ECG using machine learning techniques," *Sensors*, vol. 18, no. 4, article 1160, 2018.
- [34] H. Lin, W. Xu, N. Guan, D. Ji, Y. Wei, and W. Yi, "Noninvasive and continuous blood pressure monitoring using wearable body sensor networks," *IEEE Intelligent Systems*, vol. 33, no. 6, pp. 38–48, 2015.
- [35] F. Heydari, M. P. Ebrahim, J.-M. Redoute, K. Joe, K. Walker, and M. R. Yuce, "A chest-based continuous cuffless blood pressure method: estimation and evaluation using multiple body sensors," *Information Fusion*, vol. 54, pp. 119–127, 2019.
- [36] J. Buxi, M. Redouté, and M. R. Yuce, "A survey on signals and systems in ambulatory blood pressure monitoring using pulse transit time," *Physiological Measurement*, vol. 36, no. 3, pp. R1–R26, 2015.
- [37] Y. Z. Yoon, J. M. Kang, Y. Kwon et al., "Cuff-less blood pressure estimation using pulse waveform analysis and pulse arrival time," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1068–1074, 2018.

Research Article

n -th Order Sensor Output to Control k -DoF Serial Robot Arms

Sergio Alvarez-Rodríguez ¹ and Francisco Gerardo Peña Lecona ²

¹Tecnológico Nacional de México/Instituto Tecnológico José Mario Molina Pasquel y Henríquez/Lagos de Moreno, Mexico

²Universidad de Guadalajara, CULagos, Mexico

Correspondence should be addressed to Francisco Gerardo Peña Lecona; plecona@culagos.udg.mx

Received 2 July 2020; Revised 4 September 2020; Accepted 14 December 2020; Published 11 January 2021

Academic Editor: Antonio Fernández-Caballero

Copyright © 2021 Sergio Alvarez-Rodríguez and Francisco Gerardo Peña Lecona. 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.

Currently, zero-order sensors are commonly used as positioning feedback for the closed-loop control in robotics; thus, in order to expand robots' control alternatives, other paths in sensing should be investigated more deeply. Conditions under which the n -th order sensor output can be used to control k -DoF serial robot arms are formally studied in this work. In obtaining the mentioned control conditions, the Pickard-Lindeloff theorem has been used to prove the existence and uniqueness of the robot's mathematical model solution with n order sensory systems included. To verify that the given conditions and claims guarantee controllability for both continuous-based and variable structure-based systems, two types of control strategies are used in obtaining simulation results: the conventional PID control and a second-order Sliding Mode control.

1. Introduction

Currently, sensors of order zero are the most extensively used positioning detectors in closed-loop control systems in industrial robotics. As a matter of fact, nowadays, robotic systems nearly exclusively use incremental and absolute encoders, resolvers, and high precision potentiometers. This is due to the fact that robotic controllers require delay-free information to achieve the desired control goals: accuracy, precision, repetitiveness, avoidance of tracking error, reduction of speed error, and so on.

An extended paradigm considers that positioning sensors of higher order than zero are not reliable at all. This consideration demotivates engineers for designing sensors of order $n > 0$. Nevertheless, the authors of this work consider that the theme should be sufficiently investigated to eliminate restrictive considerations which could lead in the future, to improve sensor systems in robotics.

In this work, the performance of a k -DoF serial robot arm with dynamical inclusion of linear n -order sensors is investigated, showing that robot's properties with linear n -order sensors inclusion are invariant with respect to robot's theo-

retical dynamics, provided that the solutions of the considered linear n -order sensors exist and are unique.

Interest in studying the use of higher-order sensor devices in the robotic industry has increased due to the need to implement more efficient, faster, and optimal control systems in both types of sensor systems concentrated in a robotic arm or systems with distributed sensors connected in a network and which may also be subject to adverse environmental conditions. However, the use of higher-order sensor systems inevitably introduces time delays in robotic control systems, including those delays from the controller to the actuator and the delays from the sensor to the controller.

On the other hand, some control methodologies have been proposed, ranging from sliding mode and adaptive techniques to manage uncertainties related to the knowledge of external dynamics and disturbances [1, 2]. However, these approaches always work well in lag-free environments.

To try to compensate for these delays, low pass filter-based mechanisms have been implemented to obtain high-quality signals and be able to build control loops for control mechanisms [3].

Mathematical scaffolding to ensure unique solutions in robotic device control algorithms including higher-order sensor systems is not widely studied.

Even more, alternative solutions to deal with nonlinear systems with unmeasurable states have been proposed, in [4] where an adaptive fuzzy output-feedback control scheme is designed and in [5] where a fuzzy logic system-based switched observer is constructed to approximate the unmeasurable states.

1.1. Motivation. Sensors can be designed using a very wide variety of physical principles (e.g., inductive, capacitive); however, many phenomena must be modelled using the n -order differential equations, making them unreliable in the world of robotics.

Motivated by the fact that currently sensory systems of order $n > 0$ are not included either in industrial or in other areas of robotics, the authors are interested in expanding the study of the control of serial mechanisms using sensors that may have the potential to improve in some way the current ones.

As such, the first question is: under which conditions a serial robot mechanism can be controlled by positioning sensors of order different from zero?

If this first question is satisfactorily solved, a second one comes to mind: how can we take advantage of the sensor output for control purposes?

1.2. Contribution of This Work. The main contribution of this work is to study conditions to make possible the control of serial robot arms, taking as the feedback for the closed-loop control the output of the n -th order sensory systems. Also, a technique to obtain the mentioned feedback signal from the sensor's output is proposed.

As formal proofs for claims presented in the text are provided, the methodology of this work is rigorous.

As a proof of concept, an application example is focused on the trajectory tracking control problem (for simplicity); however, using the first time derivative of the position, the robot's joint speed control follows the same rules given herein.

1.3. Problem Statement. Figure 1 shows a block control diagram with a negative feedback summing a reference signal (Ref) to get the tracking error (e) as the input for the control block which is the controller for the system Robot-Sensors. In Figure 1, τ is the control torque, q represents the actual robot's position identified by q^* (i.e., $q^* \rightarrow q$), while \tilde{q} represents the sensor's output. If sensors are of order zero, i.e., $\bar{\beta} \tilde{q} = q$, where matrix $\bar{\beta}^{-1}$ includes the sensor's parameters of sensitiveness, the feedback of states can be taken directly from the sensor's output. Nevertheless, sensors of order n can be designed and/or included to control robotic systems. In such a case, the problem consists of recovering q from the information \tilde{q} provided by sensors, using a processing method for this task. In the following section, this problem is addressed formally.

The remainder of the document is organized as follows: in Section 2, conditions and assumptions to control serial

robots using the output of n order sensory systems are given, and formal proofs for the given claims are presented; in Section 3 an implementation example for a 3-DoF serial robot arm is performed, along with simulation results; and finally, concluding remarks are given in Section 4.

2. Robot's Model with Sensor System Inclusion

Let us consider the dynamic differential equation

$$\bar{\beta}_n \frac{d^n}{dt^n} \tilde{q} + \dots + \bar{\beta}_2 \frac{d^2}{dt^2} \tilde{q} + \bar{\beta}_1 \frac{d}{dt} \tilde{q} + \bar{\beta}_0 \tilde{q} = q - F(x, y, z, t), \quad (1)$$

where $\bar{\beta}_i \in \mathbb{R}^{k \times k}$ with $i = 0, 1, \dots, n$ represents the functional diagonal matrices, i.e., each element of $\bar{\beta}$ denoted as $\beta_{\text{line} \times \text{column}} = 0$ for all line \neq column; $\tilde{q} \in \mathbb{R}^k$ is the sensor's response vector; $q \in \mathbb{R}^k$ is the vector for the actual positions of robot joints; $F(x, y, z, t) \in \mathbb{R}^k$ represents a position and/or time dependent possible function; and k is the number of DoF of the robot.

For the case of linear sensors, (1) becomes

$$\sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} = q. \quad (2)$$

Now, let us consider the following nonlinear robot dynamics

$$\dot{x} = f(t, x), \text{ with } x(t_0) = x_0, \quad (3)$$

where vector x represents the states of the electromechanical plant, i.e., $x = [\chi_1, \chi_2]^T$; vector $f(t, x)$ is a time and/or states-possible dependent function; and x_0 is the initial condition value at $t = t_0$.

The inclusion of (2) in (3) is studied in this work, in order to obtain control conditions, according to the following claim:

Claim 1. It is possible the control of serial robots using positioning sensors of the n -th order for the closed-loop feedback of the states if the mathematical structure of the solution \tilde{q} of the differential equations that model sensor's behaviour (2) exists and is unique for all $t \geq t_0$, under the following assumptions:

Assumption 1. Sensors provide the dynamic solution of (2) for all $t \geq t_0$, i.e., it is assumed that the value for the signal \tilde{q} is available to be used at time $t \geq t_0$ by the closed-loop control system.

Assumption 2. It is assumed that the diagonal matrices $\bar{\beta}_i$ with $i = 1, \dots, n$ in (2) are known.

2.1. Existence and Uniqueness. Herein, the existence and uniqueness of the solution for the robot's dynamics are

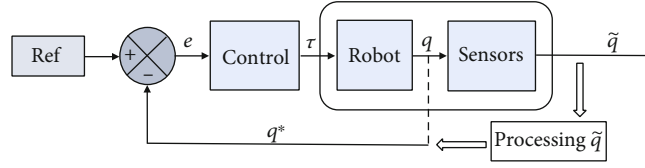


FIGURE 1: Block diagram of the closed-loop control with sensors inclusion.

verified for the case in which the solution of the sensor's dynamics exists and is unique.

Definition 1 (Lipschitz). The function $f : D \subset \mathbb{R}^{r+1} \mapsto \mathbb{R}^r$ is locally Lipschitz if it is continuous, and for all subset $\kappa \subset D$ compact exists $L \in \mathbb{R}$ such that if (x, t) and $(y, t) \in \kappa$, then,

$$\|f(x, t) - f(y, t)\| \leq L \|x - y\|, \quad (4)$$

where L is called the Lipschitz constant. (See [6] among many others.)

Theorem 1 (Picard-Lindeloff approach). For function $f : D \subset \mathbb{R}^{r+1} \mapsto \mathbb{R}^r$ locally Lipschitz, and also for the set

$$R = \{(t, x) : t_0 \leq t \leq t_0 + a, |x - x_0| \leq b\} \subset D, \quad (5)$$

let H be a bound of $\|f(t, x)\|$ in R , L the Lipschitz constant of $f(t, x)$ in R , and $\alpha = \min\{a, b/H, 1/L\}$. Then, problem (6) with initial conditions has a solution, and it is unique in the interval $[t_0, t_0 + \alpha]$ (in the Picard-Lindeloff approach).

$$\dot{\chi}_1 = f_1(t, \chi_1, \chi_2),$$

$$\dot{\chi}_2 = f_2(t, \chi_1, \chi_2),$$

$$\chi_1(t_0) = (\chi_1)_0, \chi_2(t_0) = (\chi_2)_0, \quad (6)$$

where $f = [f_1, f_2]^T$, $x = [\chi_1, \chi_2]^T$, and $(\chi_1)_0$ and $(\chi_2)_0$ are the initial values for position and velocity, respectively, at $t = t_0$.

Proof. It is assumed that \tilde{q} exists, then, also the set

$$\mathfrak{F}_0 = \left\{ \tilde{q}(t) \in \mathfrak{C}([t_0, t_0 + \alpha], \mathbb{R}^k) : \tilde{q}(t_0) = x_0, \|\tilde{q}(t) - x_0\| \leq b_0 \right\}, \quad (7)$$

exists, where $\mathfrak{C}([t_0, t_0 + \alpha], \mathbb{R}^k)$ is a Banach space with uniform norm

$$\|\tilde{q}\| = \sup_{t_0 \leq t \leq t_0 + \alpha} (\tilde{q}(t)), \quad (8)$$

and \mathfrak{F}_0 is closed in \mathfrak{C} .

According to (2) and control concept expressed in Figure 1, the identification process of q is

$$q^* = \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} + \varepsilon, \quad (9)$$

where $q^* \in \mathbb{R}^k$ is the result of processing \tilde{q} and $\varepsilon \in \mathbb{R}^k$ is an identification error given by the Euclidian norm $|q^* - q|$. Observe that $q^*(t_0) = \bar{\beta}_0 \tilde{q}(t_0) + \varepsilon$, for simplicity $\bar{\beta}_0 = I$ and $\varepsilon \rightarrow 0$, then, $q^*(t_0) \rightarrow x_0$, and the problem is reduced to the existence of the set

$$\mathfrak{F} = \left\{ q^*(t) \in \mathfrak{C}([t_0, t_0 + \alpha], \mathbb{R}^k) : q^*(t_0) = x_0, \|q^*(t) - x_0\| \leq b \right\}, \quad (10)$$

as the norm $\|q^*(t) - q^*(t_0)\|$ is bounded by b .

Let us define $\Gamma : \mathfrak{F} \rightarrow \mathfrak{C}([t_0, t_0 + \alpha], \mathbb{R}^k)$ such that the image of the function $q^*(t)$ is $\Gamma q^*(t)$ defined by

$$\Gamma q^*(t) = x_0 + \int_{t_0}^t f(s, q^*(s)) ds, \quad (11)$$

where we accomplished both

$$\begin{aligned} \Gamma q^*(t_0) &= x_0 + \int_{t_0}^{t_0} f(s, q^*(s)) ds = x_0, \\ \left\| \int_{t_0}^t f(s, q^*(s)) ds \right\| &\leq H|t - t_0| \leq H\alpha \leq b. \end{aligned} \quad (12)$$

Then, Γ is a contraction defined in a closed subset of a complete metric space, which implies that there exists a unique $q^* \in \mathfrak{F}$ such that $\Gamma q^*(t) = q^*(t)$, according to the fixed-point theorem of Banach.

Thus, $q^*(t)$ exists and is a unique solution of (6) for all $t_0 \leq t \leq t_0 + \alpha$.

2.2. Dynamics Invariance. The validity of Claim 1 is conditioned to result in the transformation of the robot's dynamics when the sensor's dynamics are included. To solve this problem, the following theorem is presented:

Theorem 2. *Dynamic properties of*

$$\dot{\chi}_1 = \chi_2,$$

$$\dot{\chi}_2 = -M^{-1} [C(q, \dot{q})\dot{q} + G(q) + F_f(q, \dot{q})] + M^{-1}(q)\tau,$$

$$\chi_1(t_0) = (\chi_1)_0, \chi_2(t_0) = (\chi_2)_0, \chi_1(t_0) = (\chi_1)_0, \chi_2(t_0) = (\chi_2)_0, \quad (13)$$

which are the dynamic robot's model with actual output q , remain invariant with respect to the transformed form

$$\dot{\chi}_1 = \chi_2,$$

$$\dot{\chi}_2 = -M^{-1} \left[C(\tilde{q}, \dot{\tilde{q}}) \frac{d}{dt} \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} + G(\tilde{q}) + F_f(\tilde{q}, \dot{\tilde{q}}) \right] + M^{-1}(\tilde{q})\tau,$$

$$\chi_1(t_0) = (\chi_1)_0, \chi_2(t_0) = (\chi_2)_0, \quad (14)$$

which is the robot with the n -th order sensor inclusion, where $\tau \in \mathbb{R}^k$ is the torque vector, $M \in \mathbb{R}^{k \times k}$ is the inertia matrix, matrix $C \in \mathbb{R}^{k \times k}$ represents the Coriolis and centripetal enforces, $G \in \mathbb{R}^k$ is the gravitational vector, and $F_{f1}^r, \dots, F_{fk}^r$ can be obtained by any parametric identification method, and each of them is the sum of viscous, Coulomb, and static friction forces.

The transformation method is given by the sensor inclusion in the robot's joints, represented by

$$\bar{\beta}_i = \begin{bmatrix} \bar{f}_i & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \bar{h}_i \end{bmatrix},$$

$$\chi_1 = q = \left[\sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1, \dots, \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_k \right]^T,$$

$$\chi_2 = \dot{q} = \left[\frac{d}{dt} \sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1, \dots, \frac{d}{dt} \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_k \right]^T, \quad (15)$$

which are the sensor's gain matrices, the articular positions, and the articular velocities, respectively.

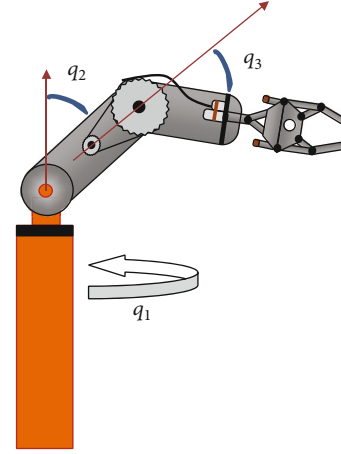


FIGURE 2: Mechanical configuration of a 3-DoF anthropomorphic robot arm: shapes of articular displacements.

Proof. The proof starts with $k = 3$, but the intermediate results are given to make evident it is the extension to $k = 2, 1$, and by induction conclude generalization to (13)–(15), *i.e.*, generalization for all $k \geq 1$.

Let us consider a serial 3-DoF anthropomorphic robot arm as defined in [7] and as illustrated by Figure 2 with positioning sensors of order n .

As dynamics is a consequence of kinematics, let us consider the form of direct and inverse kinematics with n -th order sensor inclusion.

Direct kinematics. The Denavit-Hartenberg (DH) convention (see [8, 9]) is a commonly used method for selecting frames of reference and the relationship between them to obtain the kinematics mathematical model for robotic plants. This method uses arrays as those given by (16) to represent the orientation and the positioning of the end effector (*i.e.*, the tool of the robot) [10, 11].

$$H = \begin{bmatrix} O_{3 \times 3} & X_{3 \times 1} \\ 0 & 1 \end{bmatrix}, \quad (16)$$

where $O_{3 \times 3}$ is a 3×3 matrix for the orientation, $X_{3 \times 1}$ is a 3×1 array for the position of the tool, and H is the so-called homogeneous transformation matrix (see [12]).

The DH convention claims that it is possible to obtain both position and orientation, from the $(j-1)$ -th to the j -th link, through two translation and two rotation movements, as is presented in the following homogeneous transformation matrix,

$${}^j H_i = R_{z_{j-1}, \theta_j} T_{z_{j-1}, d_j} T_{x_{j-1}, l_j} R_{x_j, \alpha_j} = \begin{bmatrix} \cos(\theta_j) & -\sin(\theta_j) \cos(\alpha_j) & \sin(\theta_j) \sin(\alpha_j) & \alpha_j \cos(\theta_j) \\ \sin(\theta_j) & \cos(\theta_j) \cos(\alpha_j) & -\cos(\theta_j) \sin(\alpha_j) & \alpha_j \sin(\theta_j) \\ 0 & \sin(\alpha_j) & \cos(\alpha_j) & d_j \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (17)$$

TABLE 1: Denavit-Hartenberg parameters.

j [DoF]	θ_j [rad]	d_j [m]	a_j [m]	α_j [rad]
1	q_1	l_1	0	$\pi/2$
2	$q_2 + \pi/2$	0	l_2	0
3	q_3	0	l_3	0

where θ , d , a , and α are the DH parameters, and their values come from specific aspects of the geometric relationship between the coordinate frames for the mechanical configuration of the robot under study and the way to select each of the coordinate frames; x_j , y_j , and z_j are the axes of the j -th coordinate frame; R_{z_{j-1},θ_j} represents the rotation matrix with respect to z_{j-1} ; T_{z_{j-1},d_j} is a translation matrix with respect to z_{j-1} ; T_{x_j,a_j} is a translation matrix with respect to x_j ; and R_{x_j,α_j} is a rotation matrix with respect to x_j (see [8–13] for a detailed procedure in obtaining (17)).

To feed equation (17) with the values of θ , d , a and α , Table 1 was constructed with the DH parameters of the robot under consideration, taking into account Figure 3 which shows the way in which the author has selected each of the coordinate frames (see [13] for more details).

Table 1 shows the articular plant positions q_1 , q_2 , and q_3 , which represent angular displacements as is shown in Figure 2; $l = [l_1, l_2, l_3]^T$ are the link lengths, which are illustrated in the right side of Figure 3.

A 3-DoF robot arm has $j = 1, 2, 3$, as such, substituting DH parameters in (17), three homogeneous transformation matrices are obtained: 0H_1 , 1H_2 , and 2H_3 . To go from system 0 to system 3, we compute ${}^0H_3 = {}^0H_1 \times {}^1H_2 \times {}^2H_3$, obtaining the final position and orientation of the tool

$${}^0H_3 = \begin{bmatrix} \eta_a & \eta_b & \sin(q_1) & \eta_e \\ \eta_c & \eta_d & -\cos(q_1) & \eta_f \\ \cos(q_2 + q_3) & -\sin(q_2 + q_3) & 0 & \eta_g \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (18)$$

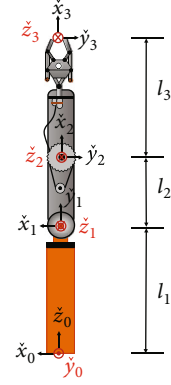


FIGURE 3: Cartesian coordinate frames and home position for the robot. Origins 1, 2, and 3 have z -axes entering the plane of drawing, and origin 0 has a y -axis leaving the plane of the drawing.

with

$$\eta_a = -\cos(q_1) \sin(q_2 + q_3),$$

$$\eta_b = -\cos(q_1) \cos(q_2 + q_3),$$

$$\eta_c = -\sin(q_1) \sin(q_2 + q_3),$$

$$\eta_d = -\sin(q_1) \cos(q_2 + q_3),$$

$$\eta_e = -\cos(q_1)(l_2 \sin(q_2) + l_3 \sin(q_2 + q_3)),$$

$$\eta_f = -\sin(q_1)(l_2 \sin(q_2) + l_3 \sin(q_2 + q_3)),$$

$$\eta_g = l_1 + l_2 \cos(q_2) + l_3 \cos(q_2 + q_3). \quad (19)$$

Thus, from the last column in (18) and including (2), direct kinematics is obtained for the robot with n -order sensor inclusion, as follows

$$X(q(t)) = [x \quad y \quad z]^T = \begin{bmatrix} -\cos\left(\sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1\right) \left(l_2 \sin\left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2\right) + l_3 \sin\left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 + \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3\right) \right) \\ -\sin\left(\sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1\right) \left(l_2 \sin\left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2\right) + l_3 \sin\left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 + \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3\right) \right) \\ l_1 + l_2 \cos\left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2\right) + l_3 \cos\left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 + \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3\right) \end{bmatrix}, \quad (20)$$

where x , y , and z are the spatial Cartesian coordinates of the end effector.

Inverse kinematics. Solving equation (20) for q_1 , q_2 , and q_3 , inverse kinematics (see [14]) is obtained. This can be done involving the Moore-Penrose pseudoinverse Jacobian matrix $J^T(JJ^T)^{-1}$ and taking into account the chain rule $\dot{X} = J\dot{q}$, for the first time derivative of q , given by $\dot{q} = J^T(JJ^T)^{-1}\dot{X}$, where J is the Jacobian matrix $J = (\partial/\partial q)X$ (see [12]). Carrying out this procedure and involving (2), the inverse kinematics for the robot is given by

$$\sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1 = \tan^{-1} \left[\frac{y}{x} \right],$$

$$\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 = \tan^{-1} \left[\frac{(l_2 + (\rho/2l_2))z^* - l_3 \sqrt{1 - (\rho/2l_2l_3)^2 \sqrt{x^2 + y^2}}}{(l_2 + (\rho/2l_2))\sqrt{x^2 + y^2} + l_3 \sqrt{1 - (\rho/2l_2l_3)^2 \cdot z^*}} \right],$$

$$\sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 = \tan^{-1} \left[\frac{2l_2l_3}{\rho} \sqrt{1 - \left(\frac{\rho}{2l_2l_3} \right)^2} \right],$$

$$\rho = x^2 + y^2 + z^{*2} - l_2^2 - l_3^2,$$

$$z^* = z - l_1, \quad (21)$$

Attitude in the orientation of the end effector. Three elements of the submatrix $O_{3 \times 3}$ inside (16) are needed to obtain ϕ , γ , and ψ which represent the roll, pitch, and yaw movements of the end effector, respectively (see [13]):

$$\begin{aligned} O_{13} &= \cos(\phi) \sin(\gamma) \cos(\psi) + \sin(\phi) \sin(\psi), \\ O_{31} &= -\sin(\gamma), \\ O_{33} &= \cos(\gamma) \cos(\psi). \end{aligned} \quad (22)$$

Now, taking (18) as the master matrix and solving for ϕ , γ , and ψ , the orientation of the end effector of 3-DoF anthropomorphic robot arms with any order linear positioning sensors is obtained as

$$\begin{bmatrix} \phi \\ \gamma \\ \psi \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1 \\ \sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 + \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \\ \pi/2 \end{bmatrix}. \quad (23)$$

The same results presented in (20)–(23) are obtained including directly in (18) the DH parameters with sensor sys-

TABLE 2: Denavit-Hartenberg parameters with sensor system inclusion.

j [DoF]	θ_j [rad]	d_j [m]	a_j [m]	α_j [rad]
1	$\sum_{i=0}^n \bar{f}_i \left(\frac{d^i}{dt^i} \right) \tilde{q}_1$	l_1	0	$\pi/2$
2	$\sum_{i=0}^n \bar{g}_i \left(\frac{d^i}{dt^i} \right) \tilde{q}_2 + \pi/2$	0	l_2	0
3	$\sum_{i=0}^n \bar{h}_i \left(\frac{d^i}{dt^i} \right) \tilde{q}_3$	0	l_3	0

tem inclusion given by Table 2, concluding that kinematics is invariant to sensor system inclusion.

2.2.1. Robot's Dynamics. Let us represent the total torque vector (τ) of the robot under consideration according to the Euler-Lagrange approach, as

$$\tau = M(q)\ddot{q} + \dot{M}(q)\dot{q} - \frac{\partial}{\partial q} \left[\frac{1}{2} \dot{q}^T M(q) \dot{q} \right] + \frac{\partial U(q)}{\partial q} + F_f^r(q, \dot{q}), \quad (24)$$

where $M \in \mathbb{R}^{3 \times 3}$ is the inertia matrix, $U \in \mathbb{R}^3$ is the potential energy vector, and $F_f^r \in \mathbb{R}^3$ is the friction force vector.

Using the methodology detailed in [15–17] to obtain the torque for the j -th link, (24) is transformed to the following form

$$\tau_i = \frac{d}{dt} \left[\frac{\partial}{\partial \dot{q}} L(q, \dot{q}) \right] - \frac{\partial}{\partial q} L(q, \dot{q}) + F_f^r(q, \dot{q}), \quad (25)$$

where $L \in \mathbb{R}^3$ is the Lagrangian given by

$$L(q, \dot{q}) = K(q, \dot{q}) - U(q), \quad (26)$$

with K and U as the kinetic and potential energies, respectively, which according to the referenced methodology are calculated for this work as

$$\begin{aligned} K(q, \dot{q}) &= \frac{1}{2} (I_1 + I_2 + I_3 + m_2 l_{c2}^2 (\sin(q_2))^2 + m_3 l_2^2 (\sin(q_2))^2 \\ &\quad + m_3 l_2 l_{c3} \sin(q_2) \sin(q_2 + q_3) + m_3 l_{c3}^2 (\sin(q_2 + q_3))^2) \dot{q}_1^2 \\ &\quad + \frac{1}{2} (I_2 + I_3 + m_2 l_{c2}^2 + m_3 l_2^2 + m_3 l_2 l_{c3} \cos(q_3)) \dot{q}_2^2 \\ &\quad + \frac{1}{2} (I_3 + m_3 l_{c3}^2) \dot{q}_3^2 + (I_2 + I_3) \dot{q}_1 \dot{q}_2 + I_3 \dot{q}_1 \dot{q}_3 \\ &\quad + (I_3 + m_3 l_2 l_{c3} \cos(q_3) + m_3 l_{c3}^2) \dot{q}_2 \dot{q}_3, \\ U(q) &= m_2 g l_{c2} [1 - \cos(\pi - q_2)] + m_3 g [(l_2 + l_{c3}) \\ &\quad - (l_2 \cos(\pi - q_2) + l_{c3} \cos(\pi - q_2 - q_3))], \end{aligned} \quad (27)$$

$[l_{c1}, l_{c2}, l_{c3}]^T$ are the inertia, mass, and length vectors, respectively, for the j -th centroid, and $g = 9.81 \text{ m/s}^2$.

Separating each element of vector (25) and calculating

$$\begin{aligned}\tau_1 &= \frac{d}{dt} \left[\frac{\partial}{\partial \dot{q}_1} L(q, \dot{q}) \right] - 0 + F_{f1}^r, \\ \tau_2 &= \frac{d}{dt} \left[\frac{\partial}{\partial \dot{q}_2} L(q, \dot{q}) \right] - \frac{\partial}{\partial q_2} L(q, \dot{q}) + F_{f2}^r, \\ \tau_3 &= \frac{d}{dt} \left[\frac{\partial}{\partial \dot{q}_3} L(q, \dot{q}) \right] - \frac{\partial}{\partial q_3} L(q, \dot{q}) + F_{f3}^r,\end{aligned}\quad (28)$$

where

$$\begin{aligned}\frac{d}{dt} \left[\frac{\partial}{\partial \dot{q}_1} L(q, \dot{q}) \right] &= [I_1 + I_2 + I_3 + m_2 l_{c2}^2 (\sin(q_2))^2 \\ &\quad + m_3 l_2^2 (\sin(q_2))^2 + m_3 l_2 l_{c3} \sin(q_2) \sin \\ &\quad \cdot (q_2 + q_3) + m_3 l_{c3}^2 (\sin(q_2 + q_3))^2] \ddot{q}_1 \\ &\quad + [2m_2 l_{c2}^2 \sin(q_2) \cos(q_2) \\ &\quad + 2m_3 l_2^2 \sin(q_2) \cos(q_2) \\ &\quad + m_3 l_2 l_{c3} \cos(q_2) \sin(q_2 + q_3) \\ &\quad + m_3 l_2 l_{c3} \sin(q_2) \cos(q_2 + q_3) \\ &\quad + 2m_3 l_{c3}^2 \sin(q_2 + q_3) \cos(q_2 + q_3)] \dot{q}_1 \dot{q}_2 \\ &\quad + [m_3 l_2 l_{c3} \sin(q_2) \cos(q_2 + q_3) \\ &\quad + 2m_3 l_{c3}^2 \sin(q_2 + q_3) \cos(q_2 + q_3)] \dot{q}_1 \dot{q}_3 \\ &\quad + [I_2 + I_3] \ddot{q}_2 + I_3 \ddot{q}_3, \\ \frac{d}{dt} \left[\frac{\partial}{\partial \dot{q}_2} L(q, \dot{q}) \right] &= [I_2 + I_3] \ddot{q}_1 + [I_2 + I_3 + m_2 l_{c2}^2 \\ &\quad + m_3 l_{c2}^2 + m_3 l_2 l_{c3} \cos(q_3)] \ddot{q}_2 \\ &\quad + [I_3 + m_3 l_{c3}^2 + m_3 l_2 l_{c3} \cos(q_3)] \ddot{q}_3 \\ &\quad - m_3 l_2 l_{c3} \sin(q_3) (\dot{q}_2 \dot{q}_3 + \dot{q}_3^2), \\ \frac{d}{dt} \left[\frac{\partial}{\partial \dot{q}_3} L(q, \dot{q}) \right] &= I_3 \ddot{q}_1 + [I_3 + m_3 l_{c3}^2 + m_3 l_2 l_{c3} \cos(q_3)] \ddot{q}_2 \\ &\quad + [I_3 + m_3 l_{c3}^2] \ddot{q}_3 - m_3 l_2 l_{c3} \sin(q_3) \dot{q}_2 \dot{q}_3, \\ \frac{\partial}{\partial q_1} L(q, \dot{q}) &= 0, \\ \frac{\partial}{\partial q_2} L(q, \dot{q}) &= [m_2 l_{c2}^2 \sin(q_2) \cos(q_2) \\ &\quad + m_3 l_2^2 \sin(q_2) \cos(q_2) \\ &\quad + \frac{1}{2} m_3 l_2 l_{c3} (\sin(q_2) \cos(q_2 + q_3) \\ &\quad + \cos(q_2) \sin(q_2 + q_3)) + m_3 l_{c3}^2 \sin \\ &\quad \cdot (q_2 + q_3) \cos(q_2 + q_3)] \dot{q}_1^2 \\ &\quad + m_2 g l_{c2} \sin(\pi - q_2) + m_3 g l_2 \sin(\pi - q_2) \\ &\quad + m_3 g l_{c3} \sin(\pi - q_2 - q_3), \\ \frac{\partial}{\partial q_3} L(q, \dot{q}) &= \left[\frac{1}{2} m_3 l_2 l_{c3} \sin(q_2) \cos(q_2 + q_3) \right. \\ &\quad + m_3 l_{c3}^2 \sin(q_2 + q_3) \cos(q_2 + q_3)] \dot{q}_1^2 \\ &\quad - m_3 l_2 l_{c3} \sin(q_3) \left(\frac{1}{2} \dot{q}_2^2 + \dot{q}_2 \dot{q}_3 \right) \\ &\quad + m_3 g l_{c3} \sin(\pi - q_2 - q_3),\end{aligned}\quad (29)$$

$$\begin{aligned}\frac{\partial}{\partial q_1} L(q, \dot{q}) &= \left[\frac{1}{2} m_3 l_2 l_{c3} \sin(q_2) \cos(q_2 + q_3) \right. \\ &\quad + m_3 l_{c3}^2 \sin(q_2 + q_3) \cos(q_2 + q_3)] \dot{q}_1^2 \\ &\quad - m_3 l_2 l_{c3} \sin(q_3) \left(\frac{1}{2} \dot{q}_2^2 + \dot{q}_2 \dot{q}_3 \right) \\ &\quad + m_3 g l_{c3} \sin(\pi - q_2 - q_3),\end{aligned}\quad (30)$$

Defining the Coriolis forces and gravitational vectors, respectively, as

$$\begin{aligned}C(q, \dot{q}) \dot{q} &= \dot{M}(q) \dot{q} - \frac{\partial}{\partial q} \left[\frac{1}{2} \dot{q}^T M(q) \dot{q} \right], \\ G(q) &= \frac{\partial U(q)}{\partial q},\end{aligned}\quad (31)$$

and including (2) in (24), the torque vector takes the form

$$\begin{aligned}\tau &= M(\tilde{q}) \frac{d^2}{dt^2} \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} + C(\tilde{q}, \dot{\tilde{q}}) \frac{d}{dt} \\ &\quad \cdot \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} + G(\tilde{q}) + F_f^r(\tilde{q}, \dot{\tilde{q}}).\end{aligned}\quad (32)$$

Rearranging results presented in (29) and (30) and also including (2), the elements of (32), ($M(\tilde{q})$, $C(\tilde{q}, \dot{\tilde{q}})$, and $G(\tilde{q})$), are obtained and given in the following form

$$M(\tilde{q}) = \begin{bmatrix} M_{11} & I_2 + I_3 & I_3 \\ I_2 + I_3 & M_{22} & M_{23} \\ I_3 & M_{32} & I_3 + m_3 l_{c3}^2 \end{bmatrix},\quad (33)$$

where

$$\begin{aligned}M_{11} &= I_1 + I_2 + I_3 + (m_2 l_{c2}^2 + m_3 l_2^2) \left(\sin \left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 \right) \right)^2 \\ &\quad + m_3 l_2 l_{c3} \sin \left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 \right) \sin \left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 + \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right) \\ &\quad + m_3 l_{c3}^2 \left(\sin \left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 + \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right) \right)^2,\end{aligned}$$

$$M_{22} = I_2 + I_3 + m_2 l_{c2}^2 + m_3 l_2 l_{c3} \cos \left(\sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right),$$

$$M_{23} = M_{32} = I_3 + m_3 l_{c3}^2 + m_3 l_2 l_{c3} \cos \left(\sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right),\quad (34)$$

$$C(\tilde{q}, \dot{\tilde{q}}) = \begin{bmatrix} 2A \frac{d}{dt} \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_2 & 0 & 2B \frac{d}{dt} \sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1 \\ -A \frac{d}{dt} \sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1 & -2D \frac{d}{dt} \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 & -2D \frac{d}{dt} \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \\ -B \frac{d}{dt} \sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1 & D \frac{d}{dt} \sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 & 0 \end{bmatrix},\quad (35)$$

where

$$A = (m_2 l_{c2}^2 + m_3 l_2^2) \sin \left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 \right) \cos \left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 \right) + \frac{m_3 l_2 l_{c3}}{2} \cos \left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 \right) \sin \left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 + \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right) + B, \quad (36)$$

$$B = \frac{1}{2} m_3 l_2 l_{c3} \sin \left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 \right) \cos \left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 + \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right) + m_3 l_{c3}^2 \sin \left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 + \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right) \cos \left(\sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 + \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right), \quad (37)$$

$$D = \frac{1}{2} m_3 l_2 l_{c3} \sin \left(\sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right), \quad (38)$$

$$G(\tilde{q}) = -g \begin{bmatrix} 0 \\ (m_2 l_{c2} + m_3 l_2) \sin \left(\pi - \sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 \right) + m_3 l_{c3} \sin \left(\pi - \sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 - \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right) \\ m_3 l_{c3} \sin \left(\pi - \sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 - \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right) \end{bmatrix}. \quad (39)$$

Additionally, the friction vector symbolic form

$$F_f^r(\tilde{q}, \dot{\tilde{q}}) = [F_{f1}^r \quad F_{f2}^r \quad F_{f3}^r]^T, \quad (40)$$

must be included in the dynamic model.

Even more, the same results presented in (33)–(39) are obtained, when (2) is directly included in the Euler-Lagrange (24) and in (31), as follows

$$\begin{aligned} \tau = & M(\tilde{q}) \left[\frac{d^2}{dt^2} \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} \right] + \dot{M}(\tilde{q}) \left[\frac{d}{dt} \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} \right] \\ & - \frac{\partial}{\partial \tilde{q}} \left[\frac{1}{2} \left[\frac{d}{dt} \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} \right]^T M(\tilde{q}) \left[\frac{d}{dt} \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} \right] \right] \\ & + \frac{\partial U(\tilde{q})}{\partial \tilde{q}} + F_f^r(\tilde{q}, \dot{\tilde{q}}), \end{aligned}$$

$$\begin{aligned} C(\tilde{q}, \dot{\tilde{q}}) & \left[\frac{d^2}{dt^2} \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} \right] \\ & = \dot{M}(\tilde{q}) \left[\frac{d}{dt} \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} \right] \\ & - \frac{\partial}{\partial \tilde{q}} \left[\frac{1}{2} \left[\frac{d}{dt} \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} \right]^T M(\tilde{q}) \right. \\ & \left. \cdot \left[\frac{d}{dt} \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} \right] \right] G(\tilde{q}) = \frac{\partial U(\tilde{q})}{\partial \tilde{q}}, \end{aligned} \quad (41)$$

where

$$\bar{\beta}_i = \begin{bmatrix} \bar{f}_i & 0 & 0 \\ 0 & \bar{g}_i & 0 \\ 0 & 0 & \bar{h}_i \end{bmatrix},$$

$$\chi_1 = q = \left[\sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1, \sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2, \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right]^T,$$

$$\chi_2 = \dot{q} = \left[\frac{d}{dt} \sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1, \frac{d}{dt} \sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2, \frac{d}{dt} \sum_{i=0}^n \bar{h}_i \frac{d^i}{dt^i} \tilde{q}_3 \right]^T, \quad (42)$$

which proves dynamics invariance to sensor inclusion.

2.2.2. Extensive Considerations. Even more, for a serial robot with 2-DoF, from Tables 1 to 2, line $i = 3$ must be removed and (16) becomes

$$H = \begin{bmatrix} O_{2 \times 2} & X_{2 \times 1} \\ 0 & 1 \end{bmatrix}, \quad (43)$$

with

$$\bar{\beta}_i = \begin{bmatrix} \bar{f}_i & 0 \\ 0 & \bar{g}_i \end{bmatrix},$$

$$\chi_1 = q = \left[\sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1, \sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 \right]^T, \quad (44)$$

$$\chi_2 = \dot{q} = \left[\frac{d}{dt} \sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1, \frac{d}{dt} \sum_{i=0}^n \bar{g}_i \frac{d^i}{dt^i} \tilde{q}_2 \right]^T.$$

Also, when a 1-DoF is considered, from Tables 1 to 2, line $i = 2, 3$ must be removed and (16) becomes

$$H = \begin{bmatrix} O_{1 \times 1} & X_{1 \times 1} \\ 0 & 1 \end{bmatrix}, \quad (45)$$

with

$$\bar{\beta}_i = [\bar{f}_i],$$

$$\chi_1 = q = \sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1, \quad (46)$$

$$\chi_2 = \dot{q} = \frac{d}{dt} \sum_{i=0}^n \bar{f}_i \frac{d^i}{dt^i} \tilde{q}_1.$$

Since procedures (16)–(40) have intermediated outcomes, it is easy to compute results for a 1-DoF and for a 2-DoF robot, to obtain the same conclusions.

Further, invoking the induction method, these results can be extended to serial robots with k -DoF for all $k \geq 1$.

In order to deal with sensor inclusion mathematical complexity, the following remarks are useful for practical implementations:

$$\frac{d^p}{dt^p} \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} = \sum_{i=0}^n \bar{\beta}_i \frac{d^{i+p}}{dt^{i+p}} \tilde{q}. \quad (47)$$

Remark 4. For the case to consider matrix β_i as a constant, the following identity should be used, in order to simplify the algebraic complications,

And for the case to consider matrix $\beta_i = I$,

$$\frac{d^p}{dt^p} \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q} = \sum_{i=p}^{n+p} \frac{d^i}{dt^i} \tilde{q}. \quad (48)$$

Remark 5. For initial conditions at the origin $\tilde{q}(0) = \dot{\tilde{q}}(0) = \ddot{\tilde{q}}(0) = \dots = 0$, the Laplace transform for sensors of the k -th order with $\beta_i = I$,

$$\mathfrak{L} \left[\sum_{i=0}^n \frac{d^i}{dt^i} \tilde{q} \right] = \tilde{Q}(s) \sum_{i=0}^n s^i = Q(s), \quad (49)$$

such that

$$f(q) = \mathfrak{L}^{-1} \left[\frac{Q(s)}{\sum_{i=0}^n s^i} \right]. \quad (50)$$

2.3. Control Using n -th Order Sensors' Output. Now, in order to take advantage of Claim 1, the following control conclusions are given:

Assumption 3. It is assumed that sensors are previously well characterized, *i.e.*, both n (the order of every sensor) and sensor's dynamics are known; this implies that $\varepsilon \rightarrow 0$.

Claim 2. Supported by Theorems 2 and 3, and if $q^* \rightarrow q$, it can be claimed that equation (2) becomes the method to identify q from \tilde{q} to control (14), which represents the robot with n -th order sensor inclusion. Thus,

$$e = X^{\text{ref}} - \sum_{i=0}^n \bar{\beta}_i \frac{d^i}{dt^i} \tilde{q}, \quad (51)$$

where $e \in \mathbb{R}^k$ is the tracking error, and $X_{\text{ref}} \in \mathbb{R}^k$ represents the reference signal.

The method is depicted in Figure 4, where the dashed squared block of the negative feedback represents (2).

Note 1. Solution \tilde{q} of (2) cannot be directly used to control (6).

Extension of Claim 2. In [18], the Lasalle-invariance principle (generally used to show asymptotic stability in the Lyapunov approach) is extended to nonautonomous switching systems. This implies that Claim 2 is valid using both continuous-based and discontinuous-based controllers, *e.g.*, Proportional Integral Derivative (PID) control and Sliding Mode (SM) controllers.

Even more, in [19], an extensive study to show the control of MIMO systems by the twisting-algorithm (a second-order SM control) is presented, showing asymptotic convergence of the system trajectories to the selected sliding manifold.

Further, in order to review the stability for biorder SM (twisting-algorithm included) control with variable PID gains, [20] can be consulted.

3. Implementation Example

To verify the validity of Claims 1 and 2, this example is designed to solve the trajectory tracking control of the anthropomorphic robot with 3-DoF (see Figures 2, 3, and 5) when n -th order sensors are included on it and when conditions (including assumptions) are accomplished.

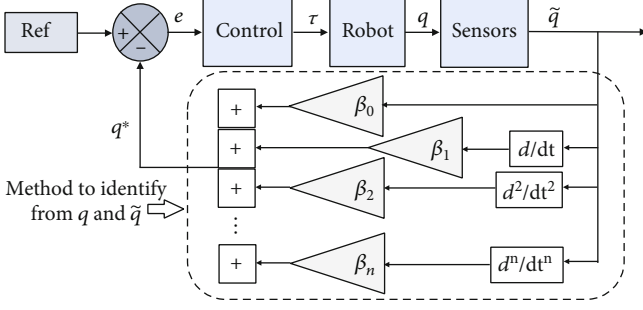


FIGURE 4: Block diagram of the closed-loop control with sensor inclusion.

Selecting $n = 2$, and

$$\begin{aligned} \bar{\beta}_0 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.9 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\ \bar{\beta}_1 &= \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.3 \end{bmatrix}, \\ \bar{\beta}_2 &= \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \end{aligned} \quad (52)$$

linear sensors, with constant coefficients of order two, zero, and one, are included in joints one, two, and three, respectively, as

$$\begin{aligned} q_1^* &= \tilde{q}_1 + \frac{1}{2} \dot{\tilde{q}}_1 + \frac{1}{10} \ddot{\tilde{q}}_1, \\ q_2^* &= \frac{9}{10} \tilde{q}_2, \\ q_3^* &= \tilde{q}_3 + \frac{3}{10} \dot{\tilde{q}}_3. \end{aligned} \quad (53)$$

Reference signal to track for the robot is selected as

$$\begin{aligned} X_1^{\text{ref}} &= \begin{cases} \frac{1}{2} & \text{if } 0 \leq t \leq 5, \\ -\frac{1}{2} & \text{if } 5 < t \leq 10, \end{cases} \\ X_2^{\text{ref}} &= \frac{2}{5} \sin\left(\frac{2}{5}\pi t\right), \\ X_3^{\text{ref}} &= \frac{2}{50}t + \frac{1}{10}. \end{aligned} \quad (54)$$



FIGURE 5: Actual 3-DoF anthropomorphic robot arm parameter model.

TABLE 3: Parameters of the actual 3-DoF anthropomorphic robot arm.

Link number i (DoF)	Total length l [m]	Centroid length l_c [m]	Total mass m [kg]	Inertia average I [$kg \cdot m^2$]
1	0.218	0.162	2.8	0.88
2	0.16	0.058	1.2	0.96
3	0.255	0.12	0.52	1.08

Thus, the ideal form of sensor systems is

$$\begin{aligned} \tilde{q}_1 + \frac{1}{2} \dot{\tilde{q}}_1 + \frac{1}{10} \ddot{\tilde{q}}_1 &= \pm \frac{1}{2}, \\ \frac{9}{10} \tilde{q}_2 &= \frac{2}{5} \sin\left(\frac{2}{5}\pi t\right), \\ \tilde{q}_3 + \frac{3}{10} \dot{\tilde{q}}_3 &= \frac{2}{50}t + \frac{1}{10}. \end{aligned} \quad (55)$$

Solving for \tilde{q} , for the open-loop system, the solution has the structure

$$\begin{aligned} \tilde{q}_1 &= \frac{1}{2} \mp \frac{1}{2} \exp\left(-\frac{5}{2}t\right) \cos\left(\frac{|\sqrt{15}|}{2}t\right) \mp \frac{5}{2|\sqrt{15}|} \exp \\ &\cdot \left(-\frac{5}{2}t\right) \sin\left(\frac{|\sqrt{15}|}{2}t\right), \end{aligned}$$

$$\begin{aligned}\tilde{q}_2 &= \frac{4}{9} \sin\left(\frac{2}{5}\pi t\right), \\ \tilde{q}_3 &= \frac{11}{125} + \frac{1}{25}t - \frac{11}{125} \exp\left(-\frac{10}{3}t\right).\end{aligned}\quad (56)$$

As the second and third terms in \tilde{q}_1 are negative for $X_1^{\text{ref}} = 1/2$ and they are positive for $X_1^{\text{ref}} = -1/2$, then, the form of solution \tilde{q} exists and is unique.

Note 2. Equations (55)–(56) are not included in the control method; they just show the ideal solution structure. For closed-loop control purposes, \tilde{q} are obtained dynamically from sensor's output.

Even more, for selecting parameters of Table 3 and friction models, the actual robot shown in Figure 5 was used, but no other application was performed using the real robot.

The following friction values are used:

$$f_1 = 0.24 \dot{q}_1 \left[kg \cdot \frac{m^2}{s} \right] + 0.16 \operatorname{sgn} \dot{q}_1 \left[kg \cdot \frac{m^2}{s^2} \right],$$

$$\begin{aligned}f_2 &= 0.21 \dot{q}_2 \left[kg \cdot \frac{m^2}{s} \right] + 0.18 \operatorname{sgn} \dot{q}_2 \left[kg \cdot \frac{m^2}{s^2} \right], \\ f_3 &= 0.17 \dot{q}_3 \left[kg \cdot \frac{m^2}{s} \right] + 0.15 \operatorname{sgn} \dot{q}_3 \left[kg \cdot \frac{m^2}{s^2} \right],\end{aligned}\quad (57)$$

where it is possible to see that if any articular velocity of the robot becomes zero, then the corresponding friction equation is reduced to zero; the sign function (sgn) appears in the second terms. This vector includes friction forces produced by ball bearings, gear boxes, and drive belts. Thus, the approximated values can be obtained directly from the manufacturer's mechanical components.

Substituting the parameters of Table 3 in (33)–(39) considering the friction model, and including sensors (52), numerical values for M , C , G , and F_f^r are given by

$$M(\tilde{q}) = \begin{bmatrix} M_{11}^* & 2.04 & 1.08 \\ 2.04 & M_{22}^* & M_{23}^* \\ 1.08 & M_{32}^* & 1.0875 \end{bmatrix}, \quad (58)$$

where

$$\begin{aligned}M_{11}^* &= 2.92 + [17.3 \sin^2(0.9\tilde{q}_2) \\ &\quad + 9.984 \sin(0.9\tilde{q}_2) \sin(0.9\tilde{q}_2 + \tilde{q}_3 + 0.3\dot{\tilde{q}}_3) \\ &\quad + 7.488 \sin^2(0.9\tilde{q}_2 + \tilde{q}_3 + 0.3\dot{\tilde{q}}_3)] \times 10^{-3}, \\ M_{22}^* &= 2.044 + [9.984 \cos(\tilde{q}_3 + 0.3\dot{\tilde{q}}_3)] \times 10^{-3}, \\ M_{23}^* &= M_{32}^* = 1.0875 + [9.984 \cos(\tilde{q}_3 + 0.3\dot{\tilde{q}}_3)] \times 10^{-3}, \\ C(\tilde{q}, \dot{\tilde{q}}) &= \begin{bmatrix} 1.8\ddot{\tilde{q}}_2 A^* & 0 & (2\dot{\tilde{q}}_1 + \ddot{\tilde{q}}_1 + 0.2\ddot{\tilde{q}}_1) B^* \\ -(\dot{\tilde{q}}_1 + 0.5\ddot{\tilde{q}}_1 + 0.1\ddot{\tilde{q}}_1) A^* & -(2\dot{\tilde{q}}_3 + 0.6\ddot{\tilde{q}}_3) D^* & -(2\dot{\tilde{q}}_3 + 0.6\ddot{\tilde{q}}_3) D^* \\ -(\dot{\tilde{q}}_1 + 0.5\ddot{\tilde{q}}_1 + 0.1\ddot{\tilde{q}}_1) B^* & \ddot{\tilde{q}}_2 D^* & 0 \end{bmatrix},\end{aligned}\quad (59)$$

with the following numerical values

$$\begin{aligned}A^* &= [17.3 \sin(0.9\tilde{q}_2) \cos(0.9\tilde{q}_2) \\ &\quad + 4.992 \cos(0.9\tilde{q}_2) \sin \\ &\quad \cdot (0.9\tilde{q}_2 + \tilde{q}_3 + 0.3\dot{\tilde{q}}_3)] \\ &\quad \times 10^{-3} + B^*,\end{aligned}\quad (60)$$

$$\begin{aligned}B^* &= [4.992 \sin(0.9\tilde{q}_2) \cos(0.9\tilde{q}_2 + \tilde{q}_3 + 0.3\dot{\tilde{q}}_3) \\ &\quad + 7.488 \sin(0.9\tilde{q}_2 + \tilde{q}_3 + 0.3\dot{\tilde{q}}_3) \cos \\ &\quad \cdot (0.9\tilde{q}_2 + \tilde{q}_3 + 0.3\dot{\tilde{q}}_3)] \times 10^{-3},\end{aligned}\quad (61)$$

$$D^* = [4.992 \sin(\tilde{q}_3 + 0.3\dot{\tilde{q}}_3)] \times 10^{-3}, \quad (62)$$

$$G(\tilde{q}) = \begin{bmatrix} 0 \\ 0.1528 \sin(\pi - 0.9\tilde{q}_2) + 0.0624 \sin(\pi - 0.9\tilde{q}_2 - \tilde{q}_3 - 0.3\dot{\tilde{q}}_3) \\ 0.0624 \sin(\pi - 0.9\tilde{q}_2 - \tilde{q}_3 - 0.3\dot{\tilde{q}}_3) \end{bmatrix}, \quad (63)$$

$$F_f^T(\tilde{q}, \dot{\tilde{q}}) = \begin{bmatrix} 0.24\dot{\tilde{q}}_1 + 0.12\ddot{\tilde{q}}_1 + 0.024\ddot{\tilde{q}}_1 + 0.16 \operatorname{sgn}(\dot{\tilde{q}}_1 + 0.5\ddot{\tilde{q}}_1 + 0.1\ddot{\tilde{q}}_1) \\ 0.189\dot{\tilde{q}}_2 + 0.18 \operatorname{sgn}(\dot{\tilde{q}}_2) \\ 0.17\dot{\tilde{q}}_3 + 0.051\ddot{\tilde{q}}_3 + 0.15 \operatorname{sgn}(\dot{\tilde{q}}_3 + 0.3\ddot{\tilde{q}}_3) \end{bmatrix}. \quad (64)$$

3.1. Controllers Used. Two techniques with different control properties are considered in this example to solve the trajectory tracking problem of the considered robot: the classical Proportional Integral Derivative (PID), and a Variable Structure System-based method.

The PID technique produces a smooth control signal, given by,

$$u_{pid} = K_p e + K_i \int_{t_0}^{t_1} e dt + K_d \frac{de}{dt}, \quad (65)$$

where $u_{pid} \in \mathbb{R}^3$ represents the value for the torque (control signal); $K_p \in \mathbb{R}^3$, $K_i \in \mathbb{R}^3$, and $K_d \in \mathbb{R}^3$ are the proportional, integral, and derivative gains, respectively; and $e = q - X^{\text{ref}} \in \mathbb{R}^3$ represents the tracking error of the control system, with $X^{\text{ref}} \in \mathbb{R}^3$ as the reference signal. For this example, the PID controller is tuned with $K_p = 1000$, $K_i = 500$, and $K_d = 400$ for every one of the robot joints.

The other considered technique is the following Sliding Mode Control (SMC) of order two,

$$u_{smc} = -r_1 \operatorname{sgn}(e) - r_2 \operatorname{sgn}(\dot{e}), \quad (66)$$

$$r_1 > r_2 \geq 0,$$

where $u_{smc} \in \mathbb{R}^3$ represents the control signal generated by the SMC; $r_1, r_2 \in \mathbb{R}^3$ are the so-called twisting parameters; and $\operatorname{sgn}(\cdot)$ is the well-known sign function. For this example, the twisting controller is tuned with $r_1 = 100$ and $r_2 = 50$, for the first joint; $r_1 = 150$ and $r_2 = 75$, for the second joint; and, $r_1 = 50$ and $r_2 = 25$, for the third joint.

Thus, the whole control system for this implementation example is depicted in Figure 6, where the reference signal (54), the control either (65) or (66), the robot (58)–(64), and the sensors (52)–(53), along with the closed-loop feedback given by Claim 2, are included.

3.2. Simulation Results. It is worth to mention that this section is not intended to show the benefits of the two control techniques presented, but to show that the robot with n -th order sensor inclusion has controllability properties.

Figures 7, 8, 9 show the performance of the first, second, and third joints of the robot in tracking a square wave form, a sine wave, and an upward sloping function, respectively (in

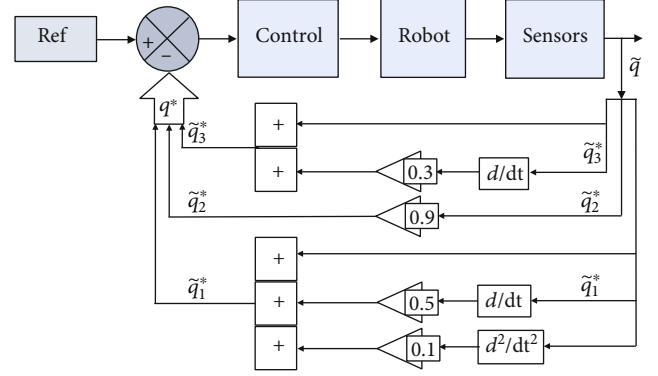


FIGURE 6: Closed-loop control diagram with sensor inclusion for this implementation example.

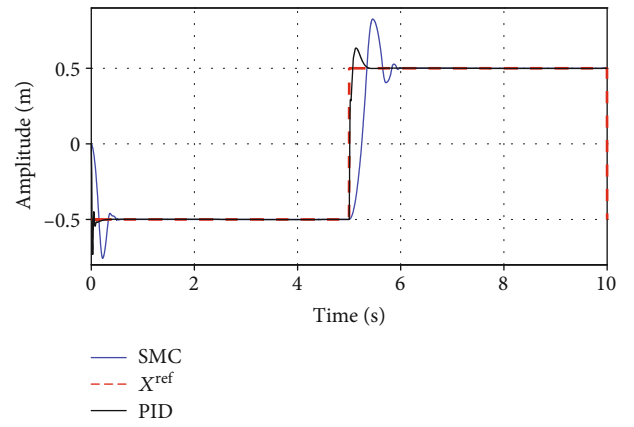


FIGURE 7: The first joint of the robot tracking a square wave as the reference signal.

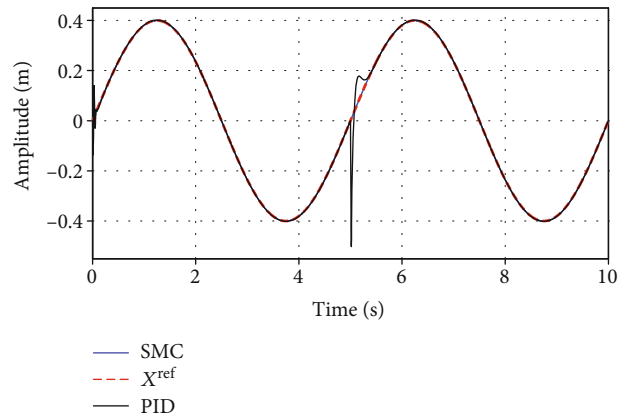


FIGURE 8: The control for the second joint of the robot in tracking a sine wave form.

dashed red line), under the control of the PID (continuous black line) and under the SMC (continuous blue line).

In these simulations, the characteristic behaviour of both controllers can be appreciated, which supports that Claim 2 is correct. The PID is tuned with high gain values to obtain fast convergence and also to avoid large overshoots. Nevertheless, the PID comes out from convergence when an adjacent joint suddenly changes, *i.e.*, the sudden change of a

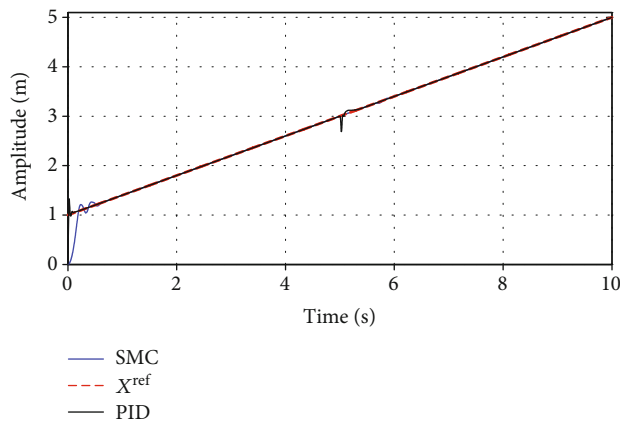


FIGURE 9: The third joint of the robot solving the trajectory tracking of an upward sloping.

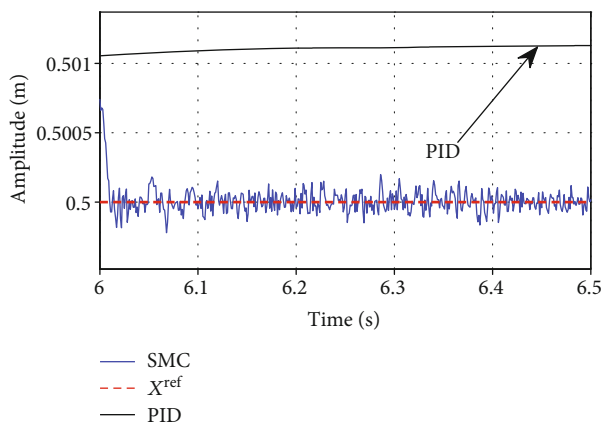


FIGURE 10: Enlarged view (zoom) of Figure 7 to show deeply the behaviour of the SMC.

joint impacts (or even shocks) the convergence of the adjacent ones. Also, it can be appreciated that the SMC is a robust control technique, because it never comes out from convergence once it is reached (after the transient state).

Even more, Figure 10 shows an enlarged view of Figure 7 from 6 to 6.5 [s], where the so-called chattering effect produced by the SMC appears.

4. Conclusions

As a result of this work, it is possible to affirm that serial robots can be controlled by taking advantage of the output of n -th order sensory systems, provided the following conditions and assumptions are fulfilled.

Conditions:

- (1) The solution of the differential equations that describe the dynamics of individual sensors must exist and be unique
- (2) The identification value of the actual robot's position should be negligible, $\varepsilon \rightarrow 0$

- (3) The dynamics of the electromechanical plant with n -th order sensor inclusion should be invariant with respect to the dynamics of the plant's output. In this work, it has been demonstrated that for the case of serial robot manipulators with k -DoF, this condition is always fulfilled

Assumptions:

- (1) The value of the plant's output \tilde{q} is always available to be used by the closed-loop control system
- (2) Coefficients of terms that describe the sensor dynamics are known. In this work, these terms are represented by diagonal matrices $\tilde{\beta}_i$ with $i = 1, \dots, n$ in (2).
- (3) Plant's dynamics is known, *i.e.*, the differential equation that model plant's dynamics is characterized

Then, serial robot manipulators are controllable using either continuous or variable structure-based controllers, and the differential equation that models sensor dynamics becomes the method to obtain the feedback for the closed-loop control.

Data Availability

The 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

Special thanks are due to Instituto Tecnológico José Mario Molina Pasquel y Henriquez, Lagos de Moreno, and Universidad de Guadalajara, CULagos, for their support to this work.

References

- [1] L. M. Capisani, T. Facchinetti, and A. Ferrara, "Real-time networked control of an industrial robot manipulator via discrete-time second-order sliding modes," *International Journal of Control*, vol. 83, no. 8, pp. 1595–1611, 2010.
- [2] S. S. Ge, Z. Li, and H. Yang, "Data driven adaptive predictive control for holonomic constrained under-actuated biped robots," *IEEE Transactions on Control Systems Technology*, vol. 20, no. 3, pp. 787–795, 2012.
- [3] Yu Kang, Zhijun Li, Xiaoqing Cao, and Dihua Zhai, "Robust control of motion/force for robotic manipulators with random time delays," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 5, pp. 1708–1718, 2013.
- [4] Y. Chang, Y. Wang, F. E. Alsaadi, and G. Zong, "Adaptive fuzzy output-feedback tracking control for switched stochastic pure-feedback nonlinear systems," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 10, pp. 1567–1582, 2019.
- [5] L. Ma, N. Xu, X. Huo, and X. Zhao, "Adaptive finite-time output-feedback control design for switched pure-feedback

- nonlinear systems with average dwell time,” *Nonlinear Analysis: Hybrid Systems*, vol. 37, p. 100908, 2020.
- [6] A. García, *Teoría de Ecuaciones Diferenciales Ordinarias*, Universidad Autónoma Metropolitana, Unidad Iztapalapa, México, 2007.
- [7] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, *Robotics: Modelling, Planning and Control*, London, Springer-Verlag, 2009.
- [8] W. S. Spong and M. Vidyasagar, *Robot Dynamics and Control*, John Wiley and Sons, New York, NY, USA, 1989.
- [9] J. Craig, *Robótica*, Pearson Educación, México, 2006.
- [10] L. Sciavicco and B. Siciliano, *Modelling and Control of Robot Manipulators*, Springer, 2005.
- [11] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot Modelling and Control*, John Wiley and Sons, New York, NY, USA, 2006.
- [12] F. Reyes-Cortés, *Robótica: Control de Robots Manipuladores*, Alfaomega, México, 2011.
- [13] F. Torres, J. Pomares, P. Gil, S. Puente, and R. Aracil, *Robots y Sistemas Sensoriales*, Prentice-Hall, Madrid, Spain, 2002.
- [14] R. Murray, Z. Li, and S. Sastry, *A Mathematical Introduction to Robotic Manipulation*, CRC Press, 1994.
- [15] F. Reyes-Cortés, *Robótica: Control de Robots Manipuladores*, México, Alfaomega, 2011.
- [16] F. Lewis, D. Dawson, and C. Abdallah, *Robot Manipulator Control: Theory and Practice*, Marcel Dekker, Inc., New York, NY, USA, 2004.
- [17] S. A. Rodríguez and C. E. Castañeda Hernández, “A dual neural network as an identifier for a robot arm,” *International Journal of Advanced Robotic Systems*, vol. 12, no. 4, p. 40, 2015.
- [18] Y. Orlov, “Extended invariance principle for nonautonomous switched systems,” *IEEE Transactions on Automatic Control*, vol. 48, no. 8, pp. 1448–1452, 2003.
- [19] A. Pisano, “On the multi-input second-order sliding mode control of nonlinear uncertain systems,” *International Journal of Robust and Nonlinear Control*, vol. 22, no. 15, pp. 1765–1778, 2012.
- [20] S. Alvarez-Rodríguez and G. Flores, “PID principles to obtain adaptive variable gains for a bi-order sliding mode control,” *International Journal of Control, Automation and Systems*, vol. 18, no. 10, pp. 2456–2467, 2020.

Research Article

Research on Innovative Trim Method for Tiltrotor Aircraft Take-Off Based on Genetic Algorithm

Xueyun Wang ^{1,2}, Jiyang Chen,¹ Qian Zhang ^{2,3}, Jingjuan Zhang,^{1,2} and Hao Cong^{1,2}

¹School of Instrument Science and Opto-Electronics Engineering, Beihang University, Beijing 100191, China

²Hefei Innovation Research Institute, Beihang University, Anhui 230012, China

³School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China

Correspondence should be addressed to Qian Zhang; zhangqian528@buaa.edu.cn

Received 3 June 2020; Revised 26 August 2020; Accepted 30 November 2020; Published 15 December 2020

Academic Editor: Antonio Fernández-Caballero

Copyright © 2020 Xueyun 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.

Tiltrotor aircraft possesses redundant actuators in take-off phase and its flight control is more complicated than ordinary aircraft because the structural and dynamic characteristics keep changing due to tilting rotors. One of the fundamental bases for flight control is trim, which provides steady flight states under various conditions and then constructs the reference trajectory. Tiltrotor aircraft trim models are described by multivariate nonlinear equations whose initial values are difficult to determine and bad initials could lead to incorrect solution for flight control. Therefore, an innovative trim method is proposed to solve this issue. Firstly, genetic algorithm (GA), which possesses strong capability in searching global optimum, is adopted to identify a coarse solution. Secondly, the coarse solution is further refined by the Levenberg-Marquardt (LM) method for precise local optimum. The innovative trim method combines the advantages of these two algorithms and is applied to a tiltrotor aircraft's flight control in the transition process of incline take-off. The limitation of trajectory is discussed, and tilt corridor is constructed. Finally, the incline take-off simulations are conducted and the effectiveness of the proposed trim method is verified through good match with the designed reference trajectory.

1. Introduction

Tiltrotor aircraft which is capable of taking-off/landing vertically and cruise at high speed simultaneously combines the advantages of rotorcraft and fixed-wing aircraft. These characteristics make tiltrotor aircraft a research hotspot for recent years [1–3]. Tilt rotors provide the aircraft with the ability of vector tension, which not only expands its operating conditions but also entangles its flight control [4, 5]. In the process of rotor tilting, it is very important yet very difficult to implement the stability control of aircraft attitudes. The adjustment of the rotor tilt angle during take-off or landing causes significant changes in the structural and mechanical properties of the aircraft, eventually resulting in obvious changes in flight control parameters. Under such circumstances, accidents are more likely to happen during take-off or landing [6–10].

To achieve attitude stabilization, trim is needed, which is applied to determine the steady flight states in the design of

the flight control algorithms. The equilibrium states of the aircraft play an important role in aircraft attitude response calculation, stability analysis, and flight control law design [11]. After trim then come the flight simulation and parameter adjustment for better flight performance [12]. However, since the tiltrotor aircraft has more actuators (control planes and propellers) than ordinary fixed-wing aircraft or rotorcraft, the aerodynamic of tiltrotor aircraft is more complex, which leads to the trim of the tiltrotor aircraft being more complicated and difficult than that of ordinary aircrafts [13]. Especially in the stage of rotor tilting when take-off or landing, the aircraft is controlled through both the direct force provided by the rotating propellers and the aerodynamic force derived from control planes. The actuator redundancy and force complexity lead to difficulty in attitude stabilization, which in turn demands more for trim. Traditional trim methods designed for ordinary aircrafts reveal the disadvantages such as lower efficiency and poor robustness. Therefore, a more efficient, robust, and accurate trim



FIGURE 1: The tiltrotor aircraft studied in this paper.

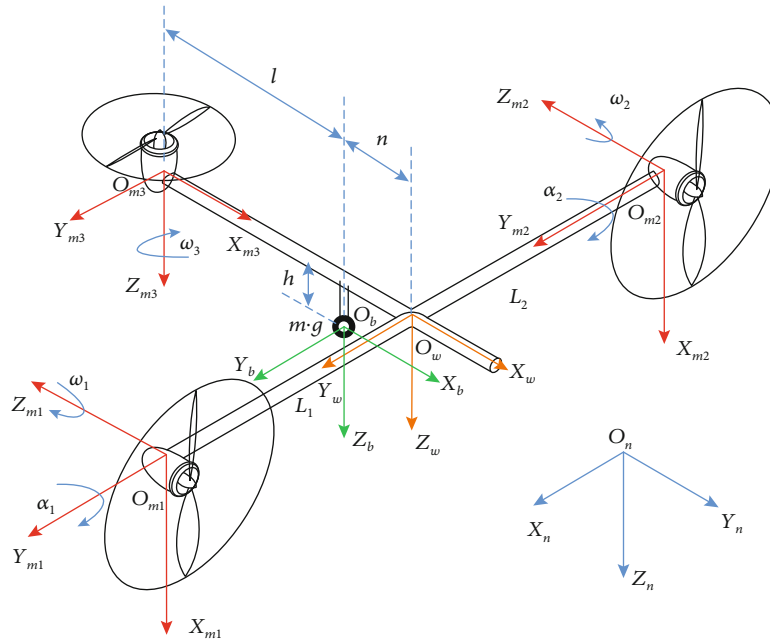


FIGURE 2: Coordinate systems of the tiltrotor aircraft.

method is needed in the design of tiltrotor aircraft fight control.

In the process of trim, it is required to solve the multivariable time-varying nonlinear equations whose initial values are difficult to determine, and the globally optimal solutions are usually not unique. Several traditional methods such as the Newton method, gradient descent method, and Levenberg-Marquardt (LM) method have been used in trim [14–16], but they are so dependent on the initial values that tiny difference may lead to the locally optimal solution rather than the globally optimal solution. Besides, inaccurate initial values increase the number of iterations or even cause divergence of the solution. Genetic algorithm (GA) is capable of searching the optimal solution effectively in global scope and achieving a robust convergent in wide range, which is suitable for solving complex nonlinear optimization problems [17, 18]. However, GA's ability of local search for optimum is weak, so the premier results derived from GA should be further promoted. They can be applied as the initial values for the traditional nonlinear programming algorithms for better local search performance. Combined with their spec-

tive advantages, the globally optimal solution for nonlinear problems can be obtained in a fast, robust, and accurate way.

In this paper, an innovative trim method for tiltrotor aircraft is proposed. Firstly, GA is adopted in the primary trim procedure to achieve a coarse solution in global scale, and secondly, the coarse solution is assigned as the initial value of the LM method for more precise local solution. In this way, GA's solution guarantees the optimality of the solution in global scale while LM's solution insures the precision of the solution in local scale. Hence, the nonlinear trim equations can be solved efficiently, robustly, and precisely. Finally, the proposed innovative trim method is verified through flight simulation of the tiltrotor aircraft under the reference take-off trajectory.

2. Mechanical Model

The tiltrotor aircraft studied in this paper, JDW-1-1, is an electric blended wing body (BWB) Unmanned Aerial Vehicle designed and manufactured by our team as shown in Figure 1. JDW-1 has several emerging technologies that make

it an advanced and unique aircraft. First, the tiltrotors give JDW-1 the ability to take off and land without a runway, the flexibility being improved. Second, the blended wing body (BWB) configuration provides good stealth and relative high lift-to-drag ratio at the same time. Finally, JDW-1 is purely electric powered without any combustion engines, so the flight control efficiency is higher and more environmentally favorable.

JDW-1 has two main propellers which can be tilted continuously by steering engines on both tips of the wings and a supplementary shrouded propeller at the tail for pitch stabilization when taking off or landing. The supplementary shrouded propeller can also be tilted along the lateral axis of the aircraft, but the range of tilt angle is only from $-\pi/6$ to $\pi/6$. Besides, it has two flaps for attitude control in level flight. Based on this structure, the mechanical model is detailed in the following aspects.

2.1. Coordinate System. The aerodynamic forces and moments are defined in airflow frame while the forces and moments of rotors are computed in rotor frame. Besides body frame and navigation frame also need to be set up in the mechanical model of six degree of freedom (DOF). A series of coordinate systems (CSs) are presented in Figure 2. The five CSs are the navigation CS $O_n - X_n Y_n Z_n$, body CS $O_b - X_b Y_b Z_b$, airflow CS $O_w - X_w Y_w Z_w$, three rotor-fixed CS $O_{mi} - X_{mi} Y_{mi} Z_{mi}$ ($i = 1, 2, 3$), and centroid CS coinciding with the body CS.

$O_n - X_n Y_n Z_n$ denotes the navigation frame which coincides with the geographic frame, north east and down (NED).

$O_b - X_b Y_b Z_b$ denotes the aircraft body frame. O_b is at the center of gravity, X_b is pointing to the front of the aircraft, Y_b is pointing right, and Z_b is defined by the right-hand rule. The rotation from NED to the body CS is defined by Euler angles.

$O_w - X_w Y_w Z_w$ denotes the air speed direction. O_w is at the center of gravity, X_w is pointing at the direction of air-speed V_a , Z_w is pointing down of the aircraft, and Y_b is defined by the right-hand rule. Rotation from the body CS to the airflow CS is defined by the angle of attack α and sideslip β .

$O_{mi} - X_{mi} Y_{mi} Z_{mi}$ ($i = 1, 2, 3$) are the rotor coordinates system which are fixed with three rotor engines, respectively, and tilt with them as shown below.

2.2. Dynamic Equations. The tiltrotor aircraft is assumed to be symmetrical relative to the plane $X_b O_b Z_b$ of the body CS while ignoring the centroid changes caused by the rotor tilting. According to Newton's second law, F being the external force acting on the airplane's center of mass and m being the mass of the aircraft, their relationship can be shown as follows:

$$F = \frac{d}{dt}(mV) = m \left(\frac{\delta V}{\delta t} + \omega \times V \right), \quad (1)$$

where $\delta V / \delta t = \dot{u}i + \dot{v}j + \dot{w}k$, where V is the aircraft speed in the body CS and u, v, w are the projection of V in the body

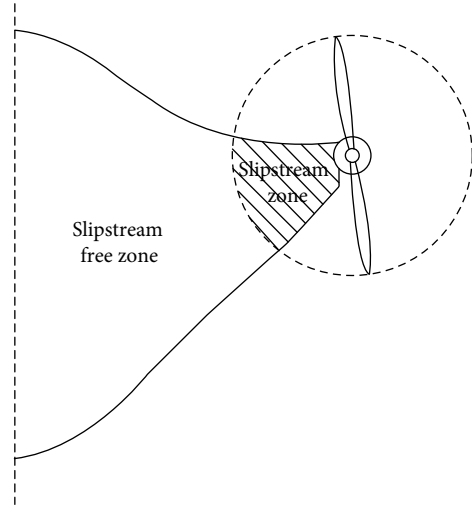


FIGURE 3: Zone division on the wing.

CS. $\omega = pi + qj + rk$ are the projections of the angular velocity of the body in the body CS. The component form is as follows:

$$\begin{cases} F_x = m(\dot{u} + qw - rv), \\ F_y = m(\dot{v} + ru - pw), \\ F_z = m(\dot{w} + pv - qu). \end{cases} \quad (2)$$

According to Euler's equation of rotation $M = I\dot{\omega} + \omega \times I\omega$, the external moment M is derived in a component form as follows:

$$\begin{cases} M_x = I_{xx}\dot{p} - (I_{yy} - I_{zz})qr - I_{xz}(\dot{r} + pq), \\ M_y = I_{yy}\dot{q} - (I_{zz} - I_{xx})pr - I_{xz}(p^2 - r^2), \\ M_z = I_{zz}\dot{r} - (I_{xx} - I_{yy})pq + I_{xz}(qr - \dot{p}). \end{cases} \quad (3)$$

Equations (2) and (3) are the six degree of freedom dynamic equations for tiltrotor aircraft.

2.3. Rotor Model. The tension and moments of the rotors expressed in body CS can be converted from rotor-fixed CS according to the following equations:

$$\begin{aligned} \begin{bmatrix} T_{rx} \\ T_{ry} \\ T_{rz} \end{bmatrix} &= \begin{bmatrix} \cos(-\alpha_r) & 0 & \sin(-\alpha_r) \\ 0 & 1 & 0 \\ -\sin(-\alpha_r) & 0 & \cos(-\alpha_r) \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ T_r \end{bmatrix}, \\ \begin{bmatrix} M_{rx} \\ M_{ry} \\ M_{rz} \end{bmatrix} &= \begin{bmatrix} 0 & -z_r & y_r \\ z_r & 0 & -x_r \\ -y_r & x_r & 0 \end{bmatrix} \begin{bmatrix} T_{rx} \\ T_{ry} \\ T_{rz} \end{bmatrix}, \end{aligned} \quad (4)$$

where T_r is the tension provided by the rotor and α_r is the tilt angle with the rotor pointing vertically being $\pi/2$ and horizontally being 0. $x_r, y_r,$ and z_r are the distances between the rotor and body center.

2.4. Wing Model. The aerodynamic forces and moments acting on the tiltrotor aircraft, such as lift force L_w , resistance D_w , lateral force C_w , rolling moment \overline{L}_A , pitching moment M_A , and yaw moment N_A in the airflow CS are calculated in the airflow CS and then converted into the body CS. Since part of the wing is under the main propeller when it is pointing vertically, the propeller's wake will aerodynamically interfere with the wing [19, 20]. Therefore, it is necessary to divide the wing area into slipstream zone and slipstream free zone to analyze the aerodynamic forces and moments, respectively. As shown in Figure 3, the slipstream zone refers to the area directly affected by the propeller's wake, while the free zone refers to the area unaffected. According to References [21, 22], the area of slipstream zone is the largest when the tiltrotor aircraft is vertically taking off and landing. When the rotor tilt angle is less than $\pi/6$, the area of the slipstream zone is 0. In practice, the slipstream area can be calculated approximately according to the following formula:

$$S_{wsr} = S_{\max} \left[\sin \left(a \left(\frac{\pi}{2} - \alpha_r \right) \right) + \cos \left(b \left(\frac{\pi}{2} - \alpha_r \right) \right) \right] \frac{u_{\max} - u}{u_{\max}}, \quad \alpha_r > \frac{\pi}{6}, \quad (5)$$

where $S_{\max} = 2\eta_{sr}R_r c_W$ is the maximum area of slipstream zone, namely, the area swept by the rotor's radius on the wing when the aircraft is hovering; η_{sr} is the slipstream correction factor; and u_{\max} is the maximum x component speed in body CS. Parameters a and b yield the following constraints and can be obtained numerically: $a = 1.386$ and $b = 3.114$.

$$\begin{cases} \sin \left(a \frac{\pi}{2} \right) + \cos \left(b \frac{\pi}{2} \right) = 1, \\ \sin \left(a \frac{\pi}{3} \right) + \cos \left(b \frac{\pi}{3} \right) = 0. \end{cases} \quad (6)$$

Assume the right wing area of the aircraft is S_{wr} while the area of slipstream-free zone is $S_{wfr} = S_{wr} - S_{wsr}$. The coordinates of the aerodynamic pressure centers of the slipstream zone and slipstream free zone on the right wing are $(x_{wsr}, y_{wsr}, z_{wsr})$ and $(x_{wfr}, y_{wfr}, z_{wfr})$, respectively. Then, the velocity of the slipstream zone and slipstream free zone at this point can be expressed as follows:

$$\begin{aligned} V_{wsr} &= V_a + \Omega_{wsr} \cdot \omega, \\ V_{wfr} &= V_a + \Omega_{wfr} \cdot \omega + \begin{bmatrix} u_{ifs} \\ 0 \\ w_{ifs} \end{bmatrix}, \end{aligned} \quad (7)$$

where the skew symmetric matrix

$$\Omega = \begin{bmatrix} 0 & z & -y \\ z & 0 & x \\ y & -x & 0 \end{bmatrix} \quad (8)$$

is defined and u_{ifs} and w_{ifs} are the decomposition of the inter-

ference of the propeller wake to the wing in the body CS. Then, the angle of attack, sideslip of slipstream zone, and slipstream free zone of the wing can be calculated as follows:

$$\begin{aligned} \alpha_{wsr} &= \sin^{-1} \left(\frac{w_{wsr}}{\sqrt{u_{wsr}^2 + w_{wsr}^2}} \right), \quad \beta_{wsr} = \sin^{-1} \left(\frac{v_{wsr}}{|V_{wsr}|} \right), \\ \alpha_{wfr} &= \sin^{-1} \left(\frac{w_{wfr}}{\sqrt{u_{wfr}^2 + w_{wfr}^2}} \right), \quad \beta_{wfr} = \sin^{-1} \left(\frac{v_{wfr}}{|V_{wfr}|} \right). \end{aligned} \quad (9)$$

Then, the force and moment generated by the right wing that determine the attitude of the aircraft are given by

$$\begin{aligned} L_{wsr} &= \frac{1}{2} \rho V_{wsr}^2 S_{wsr} C_{lwsr}, \\ D_{wsr} &= \frac{1}{2} \rho V_{wsr}^2 S_{wsr} C_{dwsr}, \\ M_{wsr} &= \frac{1}{2} \rho V_{wsr}^2 S_{wsr} C_{mwsr} l_{wr}, \\ L_{wfr} &= \frac{1}{2} \rho V_{wfr}^2 S_{wfr} C_{lwfr}, \\ D_{wfr} &= \frac{1}{2} \rho V_{wfr}^2 S_{wfr} C_{dwfr}, \\ M_{wfr} &= \frac{1}{2} \rho V_{wfr}^2 S_{wfr} C_{mwfr} l_{wr}, \end{aligned} \quad (10)$$

where l_{wr} is the mean geometric chord length of the right wing. In addition, since lateral force C_w , rolling moment \overline{L}_A , and yaw moment N_A are less affected by the wake, a

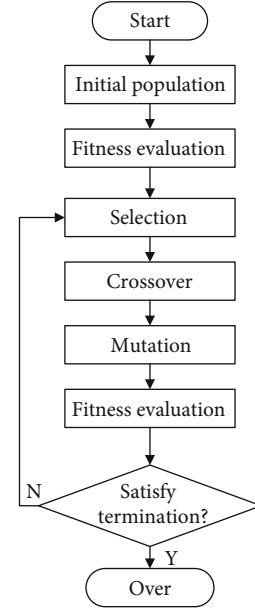


FIGURE 4: Flow chart of genetic algorithm.

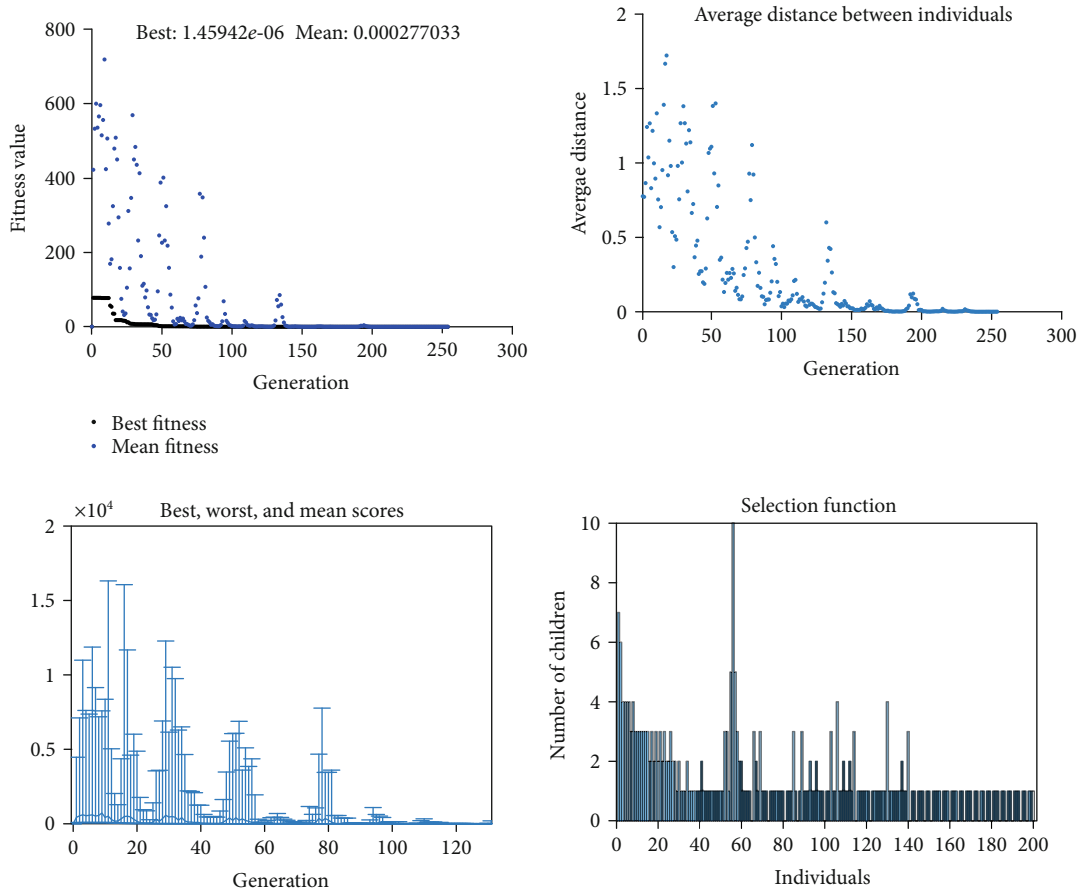


FIGURE 5: Result of GA when the tilt angle of main rotors is 50 degrees.

TABLE 1: Trim results with different size of population.

Population	T_m (N)	T_n (N)	φ (degree)	X V_a (m/s)	δ_t (degree)	δ_d (degree)	Minimum fitness score	Mean fitness score
60	-92.112	5.525	0.05521	29.273	1.3932	3.5531	6989.4	7009.1
100	-29.868	-0.2532	0.03731	32.426	0.3043	0.1151	2.06×10^{-5}	0.0615
200	-29.861	-0.2384	0.03691	32.218	0.3039	0.1140	1.52×10^{-6}	8.22×10^{-5}
500	-29.861	-0.2384	0.03691	32.218	0.3039	0.1140	2.72×10^{-6}	1.92×10^{-6}

formula without interference of the wake can be expressed as follows:

$$\begin{aligned}
 C_w &= \frac{1}{2} \rho V_w^2 S_w C_C, \\
 \overline{L}_A &= \frac{1}{2} \rho V_w^2 S_w C_l b, \\
 N_A &= \frac{1}{2} \rho V_w^2 S_w C_n b.
 \end{aligned} \tag{11}$$

Finally, the determined angle of attack and the side slip angle and the force and moment generated by the left and right wings are converted into the body CS.

2.5. *Gravity Model.* Convert the aircraft gravity into the body CS through the following equations:

$$\begin{bmatrix} T_{gx} \\ T_{gy} \\ T_{gz} \end{bmatrix} = \begin{bmatrix} -\sin \theta \\ \sin \varphi \cos \theta \\ \cos \varphi \cos \theta \end{bmatrix} mg. \tag{12}$$

3. Trim Procedure

3.1. *Genetic Algorithm.* Genetic algorithm is a random search algorithm that simulates natural selection and genetic behavior in living nature. The procedure of the genetic algorithm is

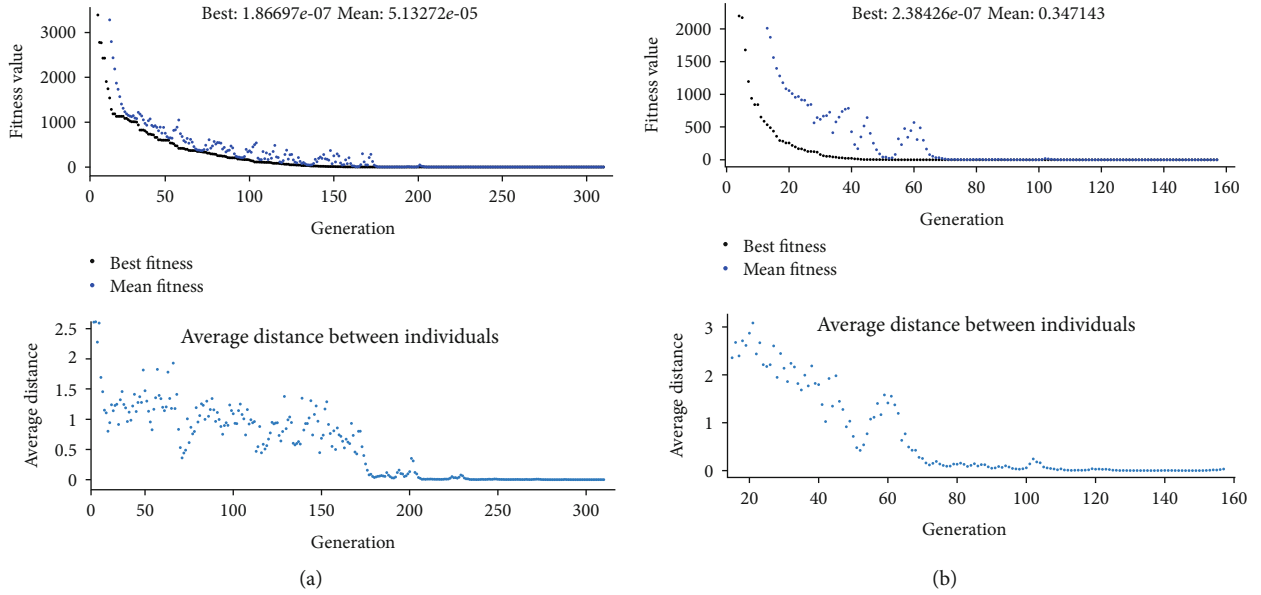


FIGURE 6: Convergence result for population size of 100 (a) and 500 (b).

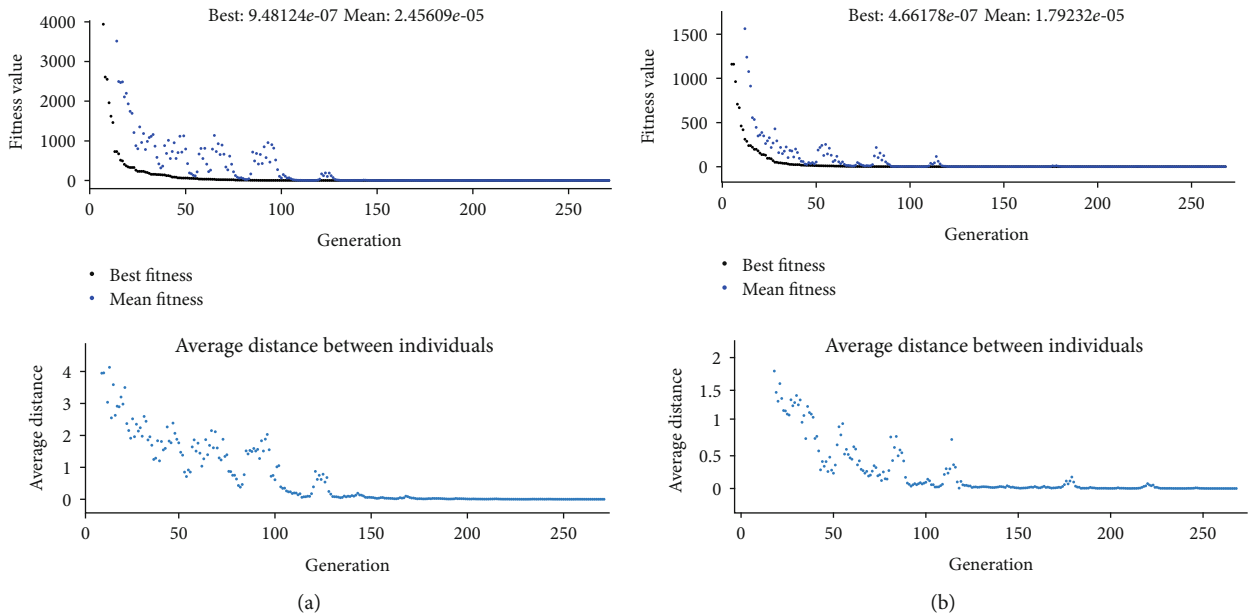


FIGURE 7: Convergence result for crossover probability of 0.2 (a) and 0.8 (b).

shown in Figure 4. The parameters to be solved are encoded to represent as chromosomes. One unsolved parameter corresponds to one chromosome, and all chromosomes constitute an individual, namely, a full state, which is a whole set of solution to equations. At the beginning, several individuals are randomly generated within the boundaries as the initial population. Then, selection, crossover, mutation, and other operations are carried out by an iterative method to exchange chromosome information. Better individuals are screened according to the fitness of the chromosomes. Finally, the chromosomes that meet the optimization target are generated.

3.2. *Trim States.* We demonstrate the trim results by taking the flight states when its rotor tilt angle is 50° as an example, since both vertical and the horizontal states are involved, making it a representative case. The trim states to be determined include the synchronous tilt angles of the main rotors $\alpha_m = \alpha_r = \alpha_l$, differential tilt angles of the main rotors $\beta_m = -\beta_r = \beta_l$, tensions of both main propellers T_l and T_r , tension and the tilt angle of the back supplementary shrouded propeller T_b and α_b , translational angle δ_t and differential angle δ_d of the aileron, and airspeed V_a and three aircraft attitudes, namely, pitch, roll, and yaw. The vector with a total of 12 states is as follows:

$$X = [\alpha_m \quad \beta_m \quad T_l \quad T_r \quad T_b \quad \alpha_b \quad \delta_t \quad \delta_d \quad V_a \quad \theta \quad \varphi \quad \psi]. \quad (13)$$

As the aircraft does not achieve enough airspeed when the tilt angle of main rotor is 50 degrees, the control efficiency of the supplementary shrouded propeller is relatively low. Therefore, it is assumed that its tension T_b and its inclination angle α_b are 0. For providing an appropriate angle of attack, assume that the aircraft's pitch is fixed at this time and the yaw is 0. For convenience, the tension of the two main rotors is expressed as the translational tension of main propellers $T_m = (T_r + T_l)/2$ and their differential tension $T_n = (T_r - T_l)/2$. For instance, if both the tilt angles of main rotors are 50 degrees, the translational tilt angle $\alpha_m = 50^\circ$ and the differential tilt angle $\alpha_n = 0$. The parameters to be trimmed can be expressed as follows:

$$X = [T_m \quad T_n \quad \varphi \quad V_a \quad \delta_t \quad \delta_d]. \quad (14)$$

As the states are the solution to the equations in the format of float real number, the encoding and decoding for the trim GA are very simple that the chromosomes are the states directly.

3.3. Trim Fitness Function. According to the dynamic equations, the forces and moments generated by each rotor when the tilt angle of main rotor is 50 degree are expressed in the body CS.

$$\begin{aligned} \begin{bmatrix} T_{ex} \\ T_{ey} \\ T_{ez} \end{bmatrix} &= \begin{bmatrix} -T_r \cos 50^\circ - T_l \cos 50^\circ \\ 0 \\ T_r \sin 50^\circ + T_l \sin 50^\circ \end{bmatrix}, \\ \begin{bmatrix} M_{ex} \\ M_{ey} \\ M_{ez} \end{bmatrix} &= \begin{bmatrix} y_r T_r \sin 50^\circ + y_l T_l \sin 50^\circ \\ -(z_r T_r + z_l T_l) \cos 50^\circ - (x_r T_r + x_l T_l) \sin 50^\circ \\ y_r T_r \cos 50^\circ + y_l T_l \cos 50^\circ \end{bmatrix}. \end{aligned} \quad (15)$$

Aerodynamic forces and moments can be expressed in the body CS as follows:

$$\begin{aligned} \begin{bmatrix} T_{wx} \\ T_{wy} \\ T_{wz} \end{bmatrix} &= \begin{bmatrix} \cos \alpha_w & 0 & -\sin \alpha_w \\ 0 & 1 & 0 \\ \sin \alpha_w & 0 & \cos \alpha_w \end{bmatrix} \begin{bmatrix} \cos \beta_w & -\sin \beta_w & 0 \\ \sin \beta_w & \cos \beta_w & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -D_w \\ C_w \\ -L_w \end{bmatrix}, \\ \begin{bmatrix} M_{wx} \\ M_{wy} \\ M_{wz} \end{bmatrix} &= \begin{bmatrix} \cos \alpha_w & 0 & -\sin \alpha_w \\ 0 & 1 & 0 \\ \sin \alpha_w & 0 & \cos \alpha_w \end{bmatrix} \begin{bmatrix} \cos \beta_w & -\sin \beta_w & 0 \\ \sin \beta_w & \cos \beta_w & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{L}_A \\ M_A \\ N_A \end{bmatrix}, \end{aligned} \quad (16)$$

where the angle of attack $\alpha_w = 2^\circ$ and sideslip angle $\beta_w = 0$.

Combined force and moment of the aircraft can be introduced as follows:

$$\begin{cases} T_x = T_{ex} + T_{wx} + T_{gx}, \\ T_y = T_{ey} + T_{wy} + T_{gy}, \\ T_z = T_{ez} + T_{wz} + T_{gz}, \\ M_x = M_{ex} + M_{wx}, \\ M_y = M_{ey} + M_{wy}, \\ M_z = M_{ez} + M_{wz}. \end{cases} \quad (17)$$

When the resultant forces and moments are zeros, the aircraft is in trim state. Therefore, the fitness function is set to be the sum of the squares of all the resultant forces and the moments as follows:

$$FitnessFcn = T_x^2 + T_y^2 + T_z^2 + M_x^2 + M_y^2 + M_z^2. \quad (18)$$

3.4. Trim Optimization Settings and Results. As the GA parameters have impact on the optimization result, they are adjusted with an experimental design method [23] and some comparisons are made here.

The GA solver of optimization toolbox in Matlab is adopted to realize the trim GA optimization. The fitness function with 6 states is established as a previous chapter describes. According to the specific situation of the tiltrotor aircraft, the lower and upper boundaries are as follows:

$$\begin{aligned} \text{Lower boundaries} &= [-500 \text{ N} \quad 500 \text{ N} \quad -10 \text{ degree} \quad 0 \text{ m/s} \quad -10 \text{ degree} \quad -10 \text{ degree}], \\ \text{Upper boundaries} &= [0 \text{ N} \quad -500 \text{ N} \quad 10 \text{ degree} \quad 50 \text{ m/s} \quad 10 \text{ degree} \quad 10 \text{ degree}]. \end{aligned} \quad (19)$$

Since fixed pitch propellers are used in the JDW-1-1, only one-direction tensions can be provided, so the translational tension of main propellers T_m is between 0 N and -500 N in the rotor coordinate system while the differential tension is between 500 N and -500 N. The roll, translational angle δ_t , and differential angle δ_d of the aileron are all constrained between -10 degree and 10 degree for flight safety.

The population size is set to 200, and the maximum limit of evolutionary algebra is 400. The crossover probability is 0.6, and the elite number is 50. The elite preservation strategy is that elites are guaranteed to survive to the next generation.

In order to remove the effect of the spread of the Fitness Function raw scores, the rank of the scores is adopted for fitness scaling rather than the score itself.

The mutation probability is not just a simple fixed ratio but a mutation function provided by Matlab. For the mutation function, adaptive feasible mode is selected, which randomly generates directions that are adaptive with respect to the last successful or unsuccessful generation. A step length is chosen along each direction so that boundaries are satisfied.

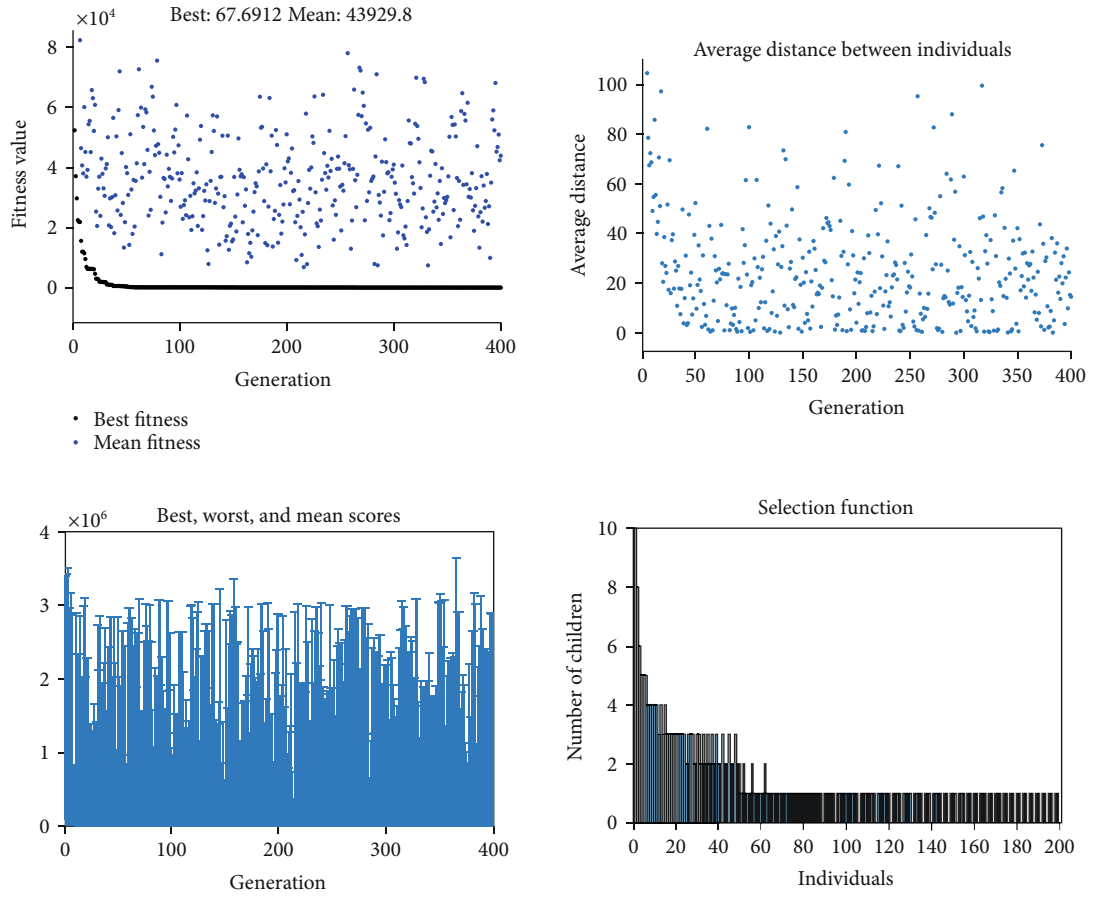


FIGURE 8: Results for mutation probability of 0.05.

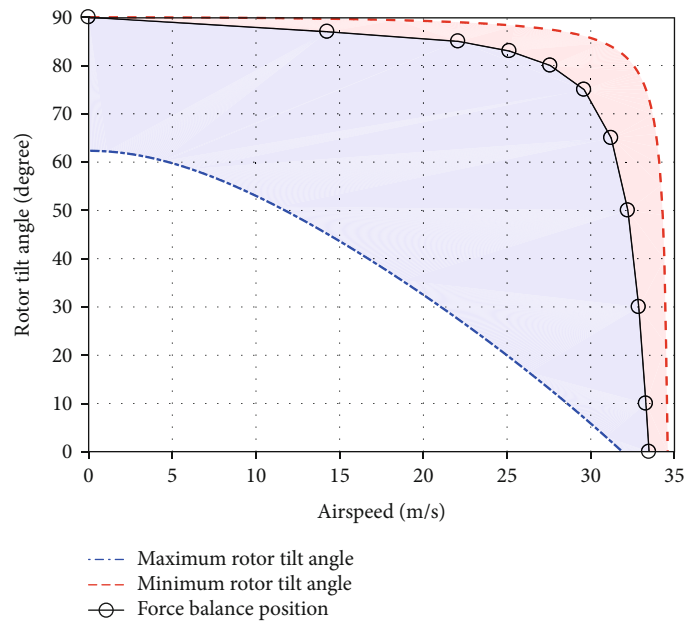


FIGURE 9: Tilt corridor of the tiltrotor aircraft.

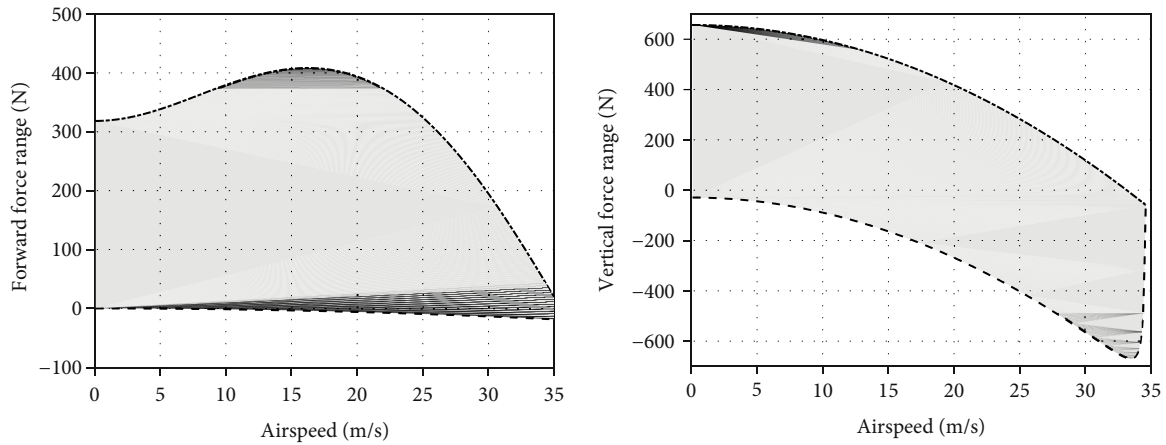


FIGURE 10: External force range of the tiltrotor aircraft.

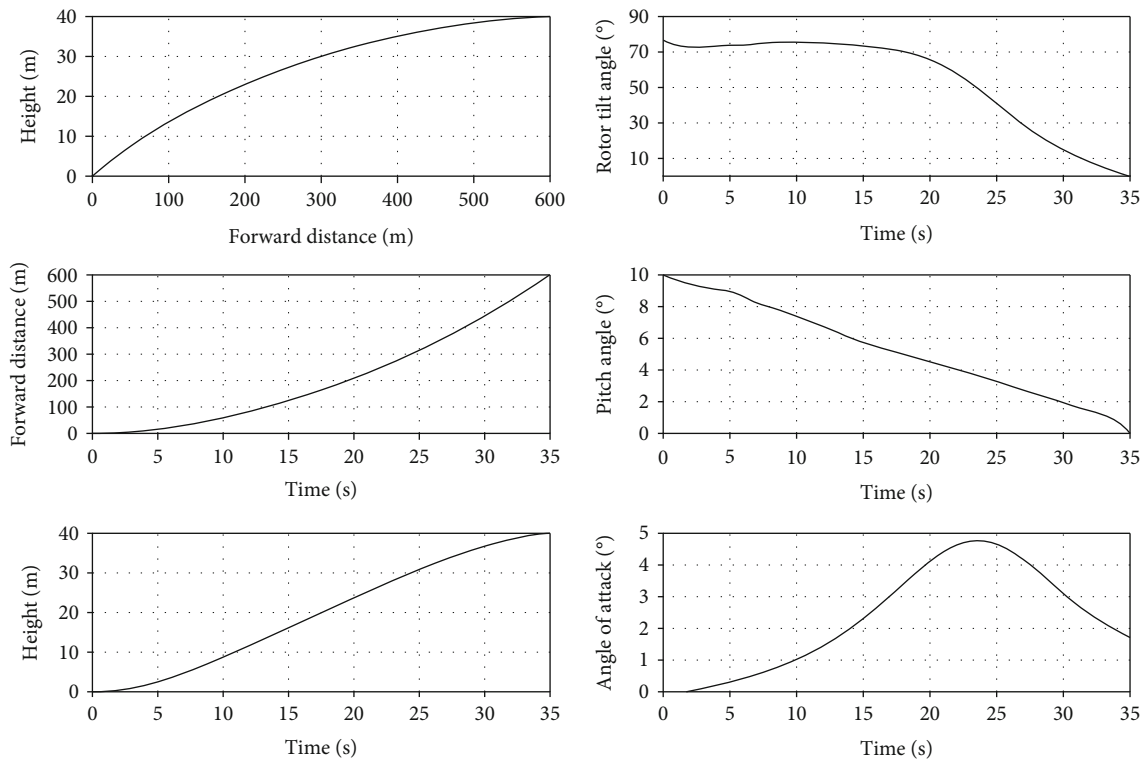


FIGURE 11: Designed aircraft take-off trajectory and attitude.

The crossover function of scattered is chosen, in which a randomly generated binary vector determines how the crossover is conducted. The crossover algorithm selects the genes where the binary vector is a 1 from the first parent and the genes where the vector is a 0 from the second parent and combines the genes to form the child. For example,

$$\begin{aligned} \text{Parent 1} &= [0.1 \ 0.2 \ 0.3 \ 0.4 \ 0.5 \ 0.6], \\ \text{Parent 2} &= [1 \ 2 \ 3 \ 4 \ 5 \ 6]. \end{aligned} \tag{20}$$

If the randomly generated binary vector is $rb = [1 \ 1 \ 0 \ 1 \ 0 \ 0]$, then, the child is $\text{child} =$

$[0.1 \ 0.2 \ 3 \ 0.4 \ 5 \ 6]$. It should be noted that individuals are randomly chosen for crossover.

For the selection function, the mode of stochastic uniform is chosen, which lays out a line in which each parent corresponds to a section of the line of length proportional to its expectation. The algorithm moves along the line in steps of equal size, one step for each parent. At each step, the algorithm allocates a parent from the section it lands on. The first step is a uniform random number less than the step size.

The GA trim results are shown in Figure 5.

According to the GA results, its fitness value continuously drops down and reaches nearly zeros after around

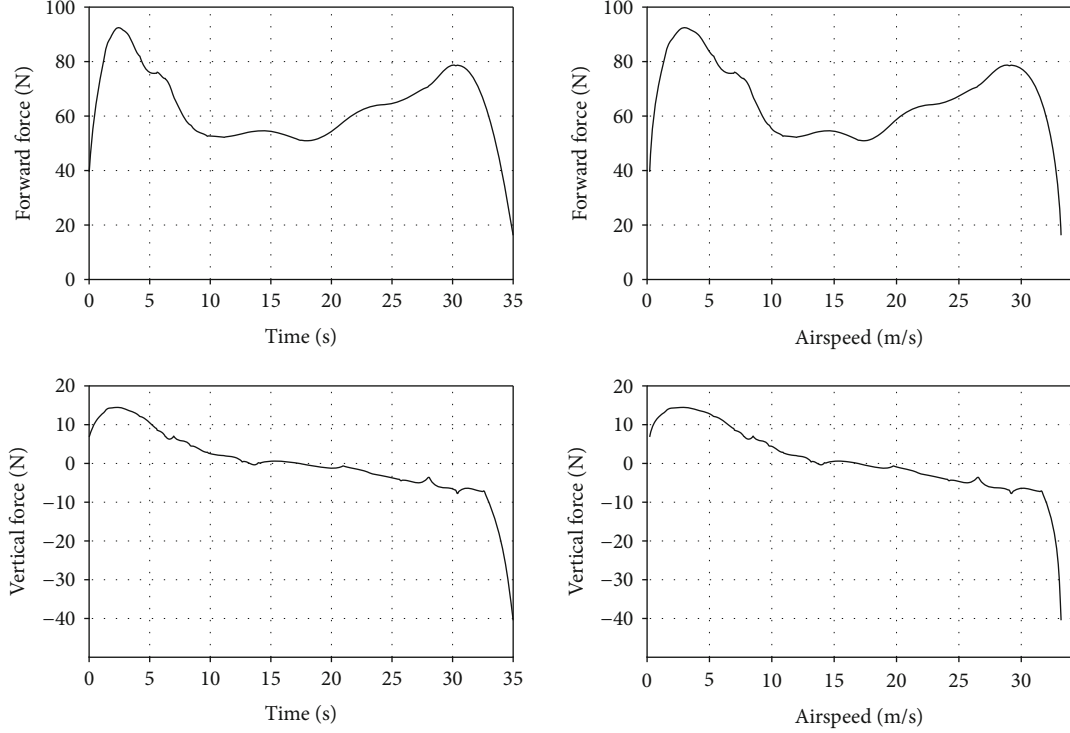


FIGURE 12: Required force for the aircraft to take off.

150 generations. Although the maximum limit of evolutionary algebra is 400, the deviation of mean fitness between two generations is less than 1×10^{-5} , when the generation reaches 262, so GA terminates and provides final optimization results.

The best individual, namely, the globally optimal solution of trim state, is as follows:

$$X_{GA}^T = \begin{bmatrix} T_m \\ T_n \\ \varphi \\ V_a \\ \delta_t \\ \delta_d \end{bmatrix} = \begin{bmatrix} -29.861 \text{ N} \\ -0.2384 \text{ N} \\ 0.0369 \text{ deg} \\ 32.218 \text{ m/s} \\ 0.3039 \text{ deg} \\ 0.1140 \text{ deg} \end{bmatrix}. \quad (21)$$

From Figure 5, the average distance between individuals reaches nearly 0 after 150 generations, indicating that the GA algorithm achieves good convergence. The worst and mean scores drop quickly in the first 60 generations, showing that the GA algorithm converges rapidly. The last subfigure shows the distribution of selection function, namely, the number of children of each individual for the second last generations. The children are generated mainly by the first 50 elites. In summary, the GA trim results achieve good convergence and succeed in providing a globally optimal solution.

Some comparisons are made with different GA parameters.

TABLE 2: Trim results of several typical moments during take-off.

A_m ($^\circ$)	T_m (N)	T_n (N)	φ (degree)	V_a (m/s)	δ_t (degree)	δ_d (degree)
15	-276.904	-0.06976	0.003355	11.56308	17.85581	-0.08725
20	-219.458	-0.14263	0.010518	16.69597	5.707258	-0.03233
30	-152.655	-0.13814	0.016089	20.82226	0.941265	0.011532
40	-115.473	-0.12574	0.019347	22.80624	-0.66647	-0.00603
50	-95.2876	-0.11508	0.023792	24.38439	-1.17095	0.075036
60	-83.6407	-0.11532	0.024298	26.18835	-1.03244	0.149144
70	-75.7513	-0.13062	0.027013	28.15163	-0.70624	0.210545
75	-72.3484	-0.13696	0.029101	29.17739	-0.51581	0.230538
80	-66.8869	-0.15094	0.03261	31.04621	-0.05611	0.233952
85	-52.493	-0.16149	0.035012	32.45665	0.194407	0.217035
90	-26.6569	-0.16494	0.036658	32.94414	0.156521	0.216939

- (1) Size of population: the size of population does not affect the optimization result much once it is more than 100. However, less than 100 populations do achieve a poor result. The trim results when the population is 60, 100, 200, and 500 are summarized in Table 1. Other parameters are fixed as previously stated. It is obvious that the differences of the final solution among the population of 100, 200, and 500 are negligible. However, the convergence speed is lower for 100 population case, as Figure 6 shows, compared with the first subfigure of Figure 5. Besides, the generation reaches 340, which also indicates poorer convergence performance for the case of 100

populations. Hence, for better accuracy and convergence considerations, the population size of 200 is selected in our research.

- (2) Crossover probability: the crossover probabilities of 0.2 and 0.8 are compared with the case of 0.6. The final optimization results are almost the same, yet the convergence performance shown in Figure 7 indicates that 0.6 is the best in the three cases.
- (3) Mutation function: mutation function has a great impact on the optimal solution. When the mutation probability is fixed to 0.05 with other settings being the same, the optimization results got much worse, which can be seen from Figure 8. Not only the fitness value keeps high, but also the convergence is poor. The final solution is not reliable. Different mutation probabilities have been testified, and the results are not satisfied. Therefore, the adaptive feasible mutation function is adopted in our research.

The premier results obtained by the GA are adopted as the initial values of the LM method to refine the trim solution in local scale, and the final trim results are obtained as follows:

$$X^T = \begin{bmatrix} T_m \\ T_n \\ \varphi \\ V_a \\ \delta_t \\ \delta_d \end{bmatrix} = \begin{bmatrix} -29.852 \text{ N} \\ -0.2231 \text{ N} \\ 0.0365 \text{ deg} \\ 32.212 \text{ m/s} \\ 0.3038 \text{ deg} \\ 0.1122 \text{ deg} \end{bmatrix}. \quad (22)$$

3.5. Tilt Corridor Design for Tiltrotor Aircraft Take-Off. Based on the innovative trim method proposed previously, the equilibrium flight states, for instance, airspeed, tensions of the rotors, and angles of flaps, can be determined under different tilting angles on a condition that proper actions are taken by the flight controller. However, the aircraft cannot always be in a “trimmed state” but also the “transition state”. Therefore, the limitations of the flight states should also be determined in order to keep the tiltrotor aircraft under control. For example, when the tiltrotor aircraft is changing the tilt angle, if the airspeed is too low but the tilt angle too small, namely, the rotor tension pointing horizontally too much, the altitude of the aircraft will decrease rapidly, and even worse, the stall will occur since the lift is not enough; On the other hand, due to the limitation of rotor performance and aircraft aerodynamic characteristics, there are maximums for the airspeed and external forces applied to the aircraft.

In order to provide limitations for the transition states and the reference for the design of the flight trajectory, it is

necessary to specify the relationship between the airspeed and the rotor tilt angle, which creates a rotor tilt angle-airspeed envelop called the tilt corridor.

According to the dynamic model of the tiltrotor aircraft, the maximum tilt angle is constrained by the z -axis force in the body CS, and the minimum tilt angle is determined by the x -axis force in the body CS. The aerodynamic force F_w and the rotor tension T are related to the airspeed. When the maximum of the rotor tension is reached, it yields to the following:

$$\begin{cases} 2T_{m \max} \cos A_{m \max} + mg + T_{b \max} + F_{wz} \leq 0, \\ T \sin A_{m \min} - F_{wx} \geq 0, \\ F_{x \max} = -2T_{m \max} \sin A_{m \max} + F_{wx}, \\ F_{z \max} = 2T_{m \max} \cos A_{m \min} + mg + F_{wz}, \end{cases} \quad (23)$$

where $T_{m \max}$ is the maximum main rotor tension, $T_{b \max}$ is the maximum back rotor tension; the $A_{m \max}$ is the maximum tilt angle; F_{wz} and F_{wx} are the aerodynamic force along the z -axis and x -axis in the body CS, respectively; and $F_{z \max}$ and $F_{x \max}$ are the maximum of the aerodynamic force along the z -axis and x -axis in the body CS, respectively.

As shown in Figure 9, the area enclosed by the dotted line and the long broken line is the feasible range of the rotor tilt angle at a specific airspeed. The solid line with circles is the equilibrium state of the aircraft calculated by the innovative trim method, in which the aircraft maintains force balance. When the tiltrotor aircraft state is within the envelope of maximum tilt angle curve and the balance curve, the aircraft is accelerating. On the contrary, in another area, the aircraft is decelerating. These constitute the transition states of the tiltrotor.

Furthermore, the external force limitation must also be considered when designing the flight path of take-off, so the range of the external forces along the x -axis (forward) and z -axis (vertical) of the tiltrotor aircraft according to airspeed are also derived and shown in Figure 10.

4. Take-Off Simulation and Verification

In this paper, an incline flight trajectory in which both height and airspeed of tiltrotor aircraft increase simultaneously in the take-off phase is designed based on the proposed trim method and the tilt corridor. The external forces, tilting angles, angle of attack, etc. with respect to time or airspeed are depicted in Figures 11 and 12.

Several typical moments during take-off are selected to fit into the trim equations in Chapter 2, in which the values of T_x and T_z in Equation (17) are external forces in perpendicular directions. It can be verified from the comparison between Figure 10 and Figure 12 that the required external forces for tiltrotor aircraft to take off according to the designed trajectory are all within the allowable range. The following results shown in Table 2 are obtained when the tiltrotor aircraft is in take-off phase.

The whole process of take-off is simulated afterwards. The trim results detailed above are applied as the target state

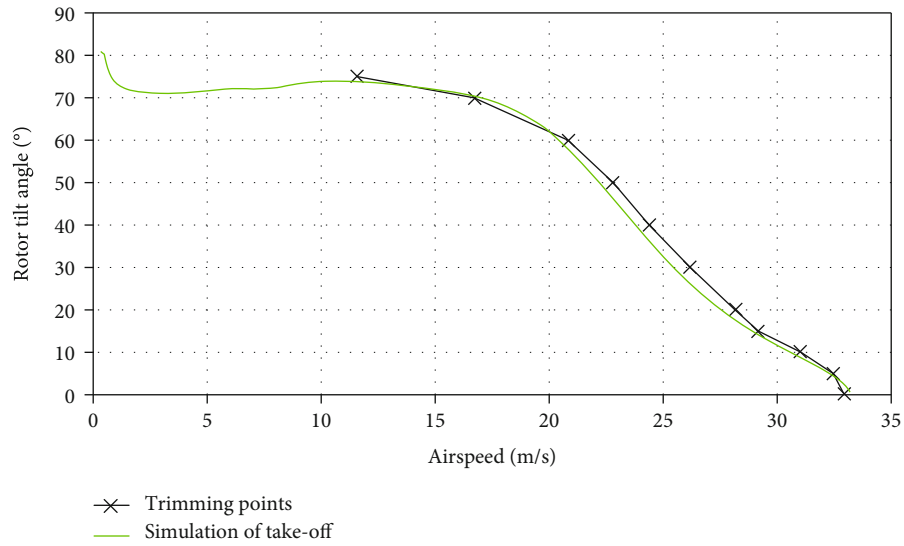


FIGURE 13: Comparison of trim results and simulation.

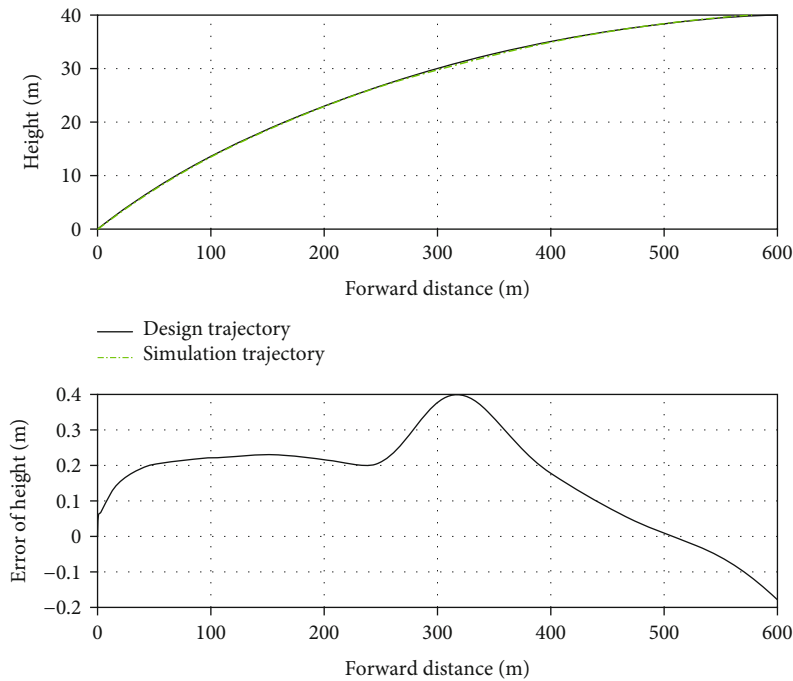


FIGURE 14: Comparison of design trajectory and simulation.

of the flight control during the take-off. The flight simulation results and the trim values are compared to be in good agreement (Figure 13), so the effectiveness of the innovative trim method is verified through flight simulation.

Besides, the simulated flight trajectory of the tiltrotor aircraft is basically consistent with the designed flight trajectory (Figure 14).

5. Conclusion

An innovative trim method for tiltrotor aircraft take-off based on a genetic algorithm is proposed in this paper.

Firstly, the genetic algorithm, which possesses strong capability in searching global optimum, is adopted to identify a coarse solution. Secondly, the coarse solution of the trim is further refined by the Levenberg-Marquardt method for precise local optimum. In addition, the innovative trim method is applied to a tiltrotor aircraft's flight control in the transition process of incline take-off. The limitation of trajectory is discussed, and the tilt corridor is constructed. Finally, the incline take-off simulations are conducted, and the effectiveness of the proposed trim method is verified through good match with the designed reference trajectory.

Data Availability

The airplane design data used to support the findings of this study have not been made available because the project involves intellectual property rights.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The research is funded by the Beijing Natural Science Foundation (4204103) and Aeronautical Science Foundation of China (2019ZC051009).

References

- [1] Z. Liu, Y. He, L. Yang, and J. Han, "Control techniques of tilt rotor unmanned aerial vehicle systems: a review," *Chinese Journal of Aeronautics*, vol. 30, no. 1, pp. 135–148, 2017.
- [2] L. Young, W. Chung, A. Paris et al., "Civil tiltrotor aircraft operations," in *11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, p. 6898, September 2011.
- [3] F. ZHANG, L. U. Ping, T. JIANG, and S. H. I. FM, *Research on modeling of the tilt tri-rotor unmanned aerial vehicle's dynamic*, DEStech Transactions on Engineering and Technology Research, 2017.
- [4] C. Papachristos, K. Alexis, and A. Tzes, "Model predictive hovering-translation control of an unmanned tri-tiltrotor robotics and automation (ICRA)," in *2013 IEEE International Conference on Robotics and Automation*, pp. 5425–5432, 2013.
- [5] B. Yuksek, A. Vuruskan, U. Ozdemir, M. A. Yukselen, and G. Inalhan, "Transition flight modeling of a fixed-wing VTOL UAV," *Journal of Intelligent and Robotic Systems*, vol. 84, no. 1-4, pp. 83–105, 2016.
- [6] J.-S. Chiou, H.-K. Tran, and S.-T. Peng, "Attitude control of a single tilt tri-rotor UAV system: dynamic modeling and each channel's nonlinear controllers design," *Mathematical Problems in Engineering*, vol. 2013, 6 pages, 2013.
- [7] F. Kendoul, I. Fantoni, and R. Lozano, "Modeling and control of a small autonomous aircraft having two tilting rotors," *IEEE Transactions on Robotics*, vol. 22, no. 6, pp. 1297–1302, 2006.
- [8] G. Notarstefano and J. Hauser, "Modeling and dynamic exploration of a tilt-rotor VTOL aircraft," *IFAC Proceedings Volumes*, vol. 43, no. 14, pp. 119–124, 2010.
- [9] D. A. Ta, I. Fantoni, and R. Lozano, "Modeling and control of a tilt tri-rotor airplane[C]. advances in computing and communications," in *2012 American control conference (ACC)*, pp. 131–136, Montreal, QC, Canada, June 2012.
- [10] J. ZHANG, L. SUN, X. QU, and L. WANG, "Time-varying linear control for tiltrotor aircraft," *Chinese Journal of Aeronautics*, vol. 31, no. 4, pp. 632–642, 2018.
- [11] W. Wang, D. Li, and C. Liu, "Helicopter flight simulation trim in the coordinated turn with the hybrid genetic algorithm," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 233, no. 3, pp. 1159–1168, 2017.
- [12] C. Papachristos and A. Tzes, "Modeling and control simulation of an unmanned tilt tri-rotor aerial vehicle[C]," in *2012 IEEE International Conference on Industrial Technology*, pp. 840–845, Athens, Greece, 2012.
- [13] T. C. Schank, *Optimal aeroelastic trim for rotorcraft with constrained, non-unique trim solutions*, Georgia Institute of Technology, 2008.
- [14] C. Friedman and O. Rand, "A highly robust trim procedure for rotorcraft simulations," in *2013 IEEE International Conference on Robotics and Automation*, p. 6361, 2008.
- [15] J. S. G. McVicar and R. Bradley, "Robust and efficient trimming algorithm for application to advanced mathematical models of rotorcraft," *Journal of Aircraft*, vol. 32, no. 2, pp. 439–442, 1995.
- [16] H. Chen, X. Wang, and T. Lei, "TiltRotor aircraft flight control study based on fuzzy logic control via genetic algorithm," in *Proceedings of 2013 Chinese Intelligent Automation Conference*, pp. 377–384, Berlin, Heidelberg, 2013.
- [17] D. A. I. Jiyang, W. U. Guohui, and Z. H. U. Guomin, "Equilibrium computation of helicopters used by hybrid genetic algorithm," *Flight Dynamics*, vol. 28, no. 1, pp. 24–28, 2010.
- [18] W. Wang and C. Liu, "GA-SM trim and validation for solving coordinated-turn state of helicopter," *Flight Dynamics*, vol. 36, no. 5, pp. 71–76, 2018.
- [19] W. Johnson, *Influence of Wake Models on Calculated Tiltrotor Aerodynamics*, National Aeronautics And Space Administration Moffett Field Ca Rotorcraft Division, 2002.
- [20] W. Tan, Y. Li, Y. Miao, and H. Lv, "The analysis of airworthiness issues influence to the development of tiltrotor aircraft," *Procedia Engineering*, vol. 80, pp. 602–608, 2014.
- [21] E. B. Carlson, Y. Zhao, and R. T. N. Chen, "Optimal tilt-rotor runway operations in one engine inoperative," AIAA-99-3961.
- [22] G. A. Ping, "Tilt-rotor aircraft modeling visualization and flight control validation," Nanjing University of Aeronautics and Astronautics, 2008.
- [23] G. Shangce, Z. Mengchu, W. Yirui, C. Jiujuun, Y. Hanaki, and W. Jiahai, *Dendritic Neuron Model with Effective Learning Algorithms for Classification, Approximation and Prediction*, vol. 30, no. 2, 2019IEEE Transactions on Neural Networks and Learning Systems, 2019.

Research Article

A Step Length Estimation Model of Coefficient Self-Determined Based on Peak-Valley Detection

Wenxia Lu , Fei Wu , Hai Zhu, and Yujin Zhang

School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, 201620, China

Correspondence should be addressed to Fei Wu; fei_wu1@163.com

Received 20 May 2020; Revised 8 July 2020; Accepted 31 October 2020; Published 27 November 2020

Academic Editor: Rafael Morales

Copyright © 2020 Wenxia Lu 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.

Without any preinstalled infrastructure, pedestrian dead reckoning (PDR) is a promising indoor positioning technology for pedestrians carrying portable devices to navigate. Step detection and step length estimation (SLE) are two essential components for the pedestrian navigation based on PDR. To solve the overcounting problem, this study proposes a peak-valley detection method, which can remove the abnormal values effectively. The current step length models mostly depend on individual parameters that need to be predetermined for different users. Based on fuzzy logic (FL), we establish a rule base that can adjust the coefficient in the Weinberg model adaptively for every detected step of various human shapes walking. Specifically, to determine the FL rule base, we collect user acceleration data from 10 volunteers walking under the combination of diverse step length and stride frequency, and each one walks 49 times at all. The experimental results demonstrate that our proposed method adapts to different kinds of persons walking at various step velocities. Peak-valley detection can achieve an average accuracy of 99.77% during 500 steps of free walking. Besides, the average errors of 5 testers are all less than 4 m per 100 m and the smallest one is 1.74 m per 100 m using our coefficient self-determined step length estimation model.

1. Introduction

Serving as a requisite part for the pedestrian of this era, Location-Based Service (LBS) has been widely provided by portable devices [1], due to the miniaturization, low energy, and low cost of Micro-Electro-Mechanical System (MEMS) sensors. Localization is the basic precondition to realize LBS in both outdoor and indoor environments. Global Positioning System (GPS) has been universally used for outdoor positioning but does not work well for indoor positioning since GPS signals can be seriously blocked and influenced by reinforced concrete buildings [2]. Due to the limitation of positioning accuracy, cost, the complexity of indoor space, and other factors, LBS for a pedestrian in the indoor environments has not been popularized [3]. Existing indoor positioning technologies include Bluetooth [4], ultrawideband (UWB) [5], and Wi-Fi [6]. Despite the great performance achieved, these methods are infrastructure-based, which require equipment predeployment. In addition, the expense of deployment is proportional to the size of the indoor area, which prevents the widespread application of these solutions.

Different from the infrastructure-based technologies, pedestrian dead reckoning (PDR) obtains the measurement and statistics of pedestrian steps, step length, and direction by using sensor data [7]. Binding the sensor to a certain part of body, several body locations have already been tested, e.g., foot [8], chest [9], legs [10], and waist [11]. To check the data characteristic of an accelerometer attached to different parts of the body in the same motion state, Wu et al. [12] compared the acceleration data where the accelerometer is bound at the feet, legs, and waist, respectively; then, the results indicated that the foot-mounted data has the strongest periodicity and regularity, while the waist gives the accelerometer a more stable measurement environment. Different from attaching sensors to a part of body directly, the position information obtained through the inertial measurement unit (IMU) built in the commodity smartphones or tablets contains a large error. The reasons are as follows: firstly, it is a problem to detect every change of pedestrian behaviour precisely, since the device is not an accessory fixed on the highly dynamic human body. Furthermore, the step length varies from person to person, even if the same individual in

different states (when walking fast, the step is larger, and when walking slowly, the step is smaller). Therefore, the accuracy of positioning based on PDR depends essentially on step detection and step length estimation (SLE).

A great many of step detection methods are mainly based on the peak detection and threshold detection [13, 14]. The peak detection method counts steps in line with the appearance of wave peak, while the threshold method detects the step according to whether the signals reach the threshold or not. However, these two ways cannot remove the noise well, which leads to the problems of overcounting and false counting. Some researches proposed features based on the acceleration to reduce the error of step detection [15–19]. In reference [15], the authors combined peak detection with four features (minimal peak distance, minimal peak prominence, dynamic thresholding, and vibration elimination) and their method detects the number of steps with an ultrahigh accuracy, but it is computationally complex. Kang et al. [16] proposed to extract frequency domain features of the three-dimensional (3D) angular velocities returned by the gyroscope through FFT (Fast Fourier Transform). However, this method is not applicable for fast walking. Gu et al. [17] come up with three features (periodicity, similarity, and continuity) considering users' false walking state. Nevertheless, it may lead to undercounting problem. Various features have been presented in the literature [18] which are split into three groups, including time domain, frequency domain, and others.

Typical SLE models are divided into four categories: constant model [20], linear models [21], nonlinear models [22–24], and neural network models [25, 26]. One of the most popular SLE models was presented by Weinberg [22], which estimates each step of a pedestrian dynamically by establishing the relationship between step length and acceleration with a constant k to adjust. But the k -value needs to be predetermined for different individuals and changes from step to step. In this regard, several researches have improved the model [27–29]. Based on the waist-mounted type, Lai et al. [27] designed a fuzzy rule base for personalized SLE with step strength, and frequency as input variables and step length as an output variable. Nevertheless, it is not flexible enough for different individuals and has no universality due to the rule base needs to be adjusted for different subjects. Ho et al. [28] derived a k -factor as a polynomial function of the average step velocity, but the estimation results contain large errors under long-distance walking. Strozzi et al. [29] developed a formula about k -value with a per-subject constant β . This method had good performance in the case of normal walking speed. However, the errors became excessively large when the subject is walking at high or low speed.

The objective of this study is to propose a peak-valley detection method that detects the pair of peak-valley to overcome the problem of overcounting. Based on the fuzzy logic (FL) algorithm, this paper designs a fuzzy controller to make the constant coefficient k in the Weinberg [22] nonlinear SLE model adjusted adaptively to each detected step, which is suitable for different kinds of people walking at various velocities. Beyond that, it is possible to estimate pedestrian

walking distance accurately by accumulating every estimated step length.

The remainder of this paper is arranged as follows: in Section 2, we describe peak-valley detection, including data collection and preprocessing. In Section 3, a coefficient self-determined step length estimation method is proposed. In Section 4, the experimental results are described in detail and we compare our proposed method with other five similar SLE algorithms. In Section 5, we discuss the deficiency of our method and the research in the future. Finally, the conclusion is presented in Section 6.

2. Peak-Valley Detection

Step detection is an issue in PDR that will cause substantial errors in the final positioning results because of overcounting, undercounting, and false detection [15]. The acceleration signal is usually used to detect the pedestrian steps. The three-axis accelerometer built in iPad Air 2 (iOS 12.1.4) is utilized for step detection in this paper. The developer can access both the raw values recorded by the hardware and a processed version of those values through the Core Motion framework [30]. In this paper, steps are calculated based on the user acceleration, a processed accelerometer value that reflects only the acceleration caused by the user and not the acceleration caused by gravity.

Figure 1 illuminates the amplitude of three-axis user acceleration collected from the iPad held in hand during normal walking and the Signal Vector Magnitude (SVM) [27]. SVM is defined as the root square of the sum of the squared acceleration signal of each axis:

$$\text{SVM}(j) = \sqrt{\text{Acc}_x(j)^2 + \text{Acc}_y(j)^2 + \text{Acc}_z(j)^2}, \quad (1)$$

where Acc_x , Acc_y , and Acc_z denote the measured user acceleration in the x -axis, y -axis, and z -axis, respectively, of the accelerometer triad at the j^{th} epoch. The unit g is the gravity acceleration.

As seen in Figure 1, the results of the walking data indicate that user acceleration contains evident noise, especially in the x -axis. User acceleration along the z -axis is more periodic, symmetric, and stable than the other two axes and SVM, which represents that the pedestrian has a more obvious periodic fluctuation in the vertical axis when walking. Therefore, based on the z -axis user acceleration, we propose to detect the pair of peak-valley to improve the accuracy of peak detection. Figure 2 shows the flow chart of peak-valley detection.

First, we obtain the z -axis user acceleration after a low-pass filter with 3 Hz cut-off frequency. The z -axis user acceleration before and after using a low-pass filter is shown in Figure 3, collected by a volunteer holding an iPad walking along a straight path, from which we can observe that the data filtered fluctuates much more smoothly. The cut-off frequency 3 Hz is determined by walking frequency range. To conclude the range of walking frequency, we set a 16 m long straight test path and ask 10 healthy volunteers (5 male and 5 female, age from 22 to 28, height from 1.58 m to 1.82 m,

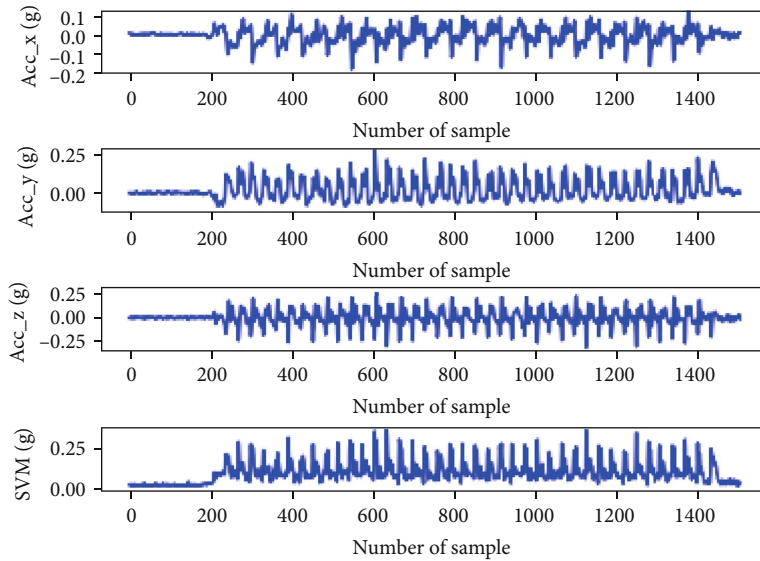


FIGURE 1: User acceleration amplitude along with the three-axis and the SVM value.

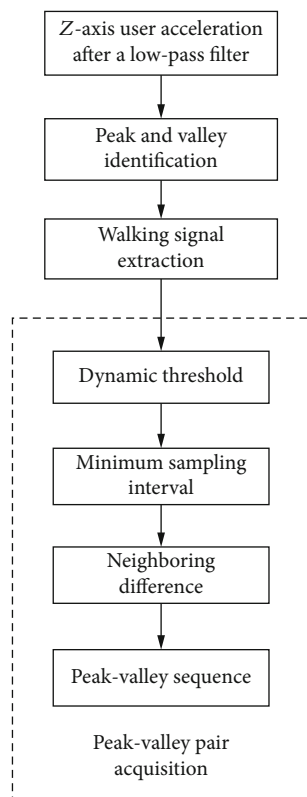


FIGURE 2: The flow chart of peak-valley detection.

weight from 47 kg to 88 kg) to walk along the path three times at different speeds (slow, normal, and fast) and count the number of steps in each time. Meanwhile, we use a stopwatch to record the walking time of every volunteer in each walk and a tape to measure the actual walking distance. Therefore, we can know that the walking frequency of pedestrians does not exceed 3 Hz. Besides, we also get the common walking step length which ranges from 0.4 m to 0.8 m.

Then, the peak and valley values both in the static and walking states are identified, while ignoring the negative peaks and positive valleys, as shown in Figure 4(a). Figure 4(b) shows the next step of extracting walking signals. Specifically, to keep the peaks and valleys in the walking state completely, the value $a_p(i)$ or $a_v(j)$ is considered to be the static one if it meets the following condition:

$$\begin{cases} \frac{\sum_i^{i+2} a_p(i)}{3} < 0.01, i = 1, 2, \dots, k-2, \\ \frac{\sum_j^{j+2} a_v(j)}{3} > -0.01, j = 1, 2, \dots, l-2, \end{cases} \quad (2)$$

where a_p and a_v are the identified peaks and valleys, i and j represent the sequence number, k and l are the number of the detected peaks and valleys after the stage of peak and valley identification, and 0.01 and -0.01 are the maximum and minimum amplitude of the z-axis user acceleration in the static. After removing most of the static values by formula (2), we mark the peaks and valleys reserved in the order as $a_{\text{peak}}(i)$ and $a_{\text{valley}}(j)$.

To obtain the paired peak-valley accurately, the following stages are executed:

(i) Dynamic threshold

The dynamic threshold is utilized to remove false values that are far from the true one. For $a_{\text{peak}}(i)$, it will be retained if it meets two conditions:

$$a_{\text{peak}}(i) \geq c * \frac{\sum_i^{i+9} a_{\text{peak}}(i)}{10} \&\& a_{\text{peak}}(i) \geq \text{Th1}. \quad (3)$$

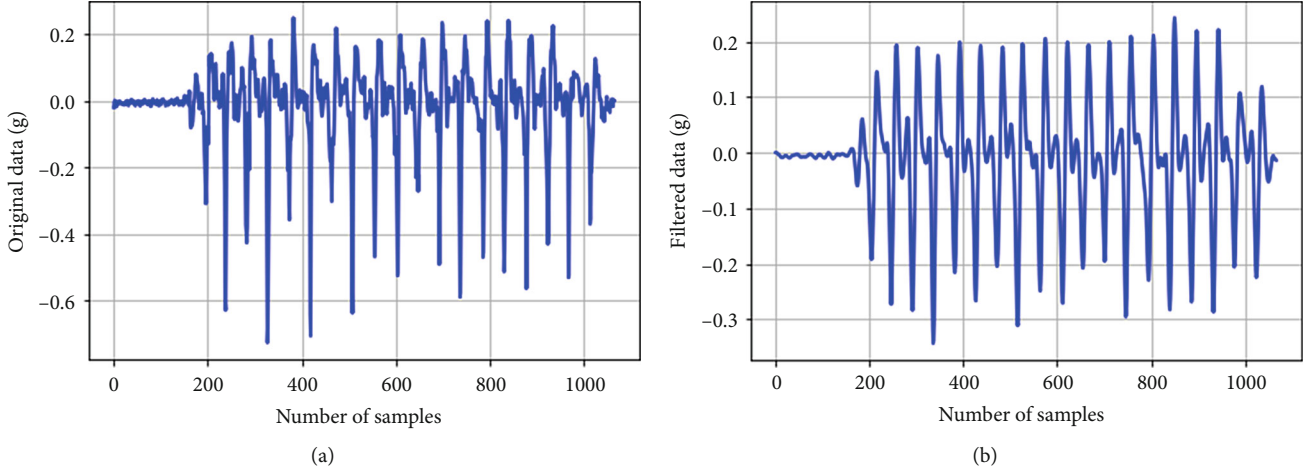


FIGURE 3: The signal before and after using a low-pass filter: (a) z -axis user acceleration before filtering; (b) z -axis user acceleration data after filtering.

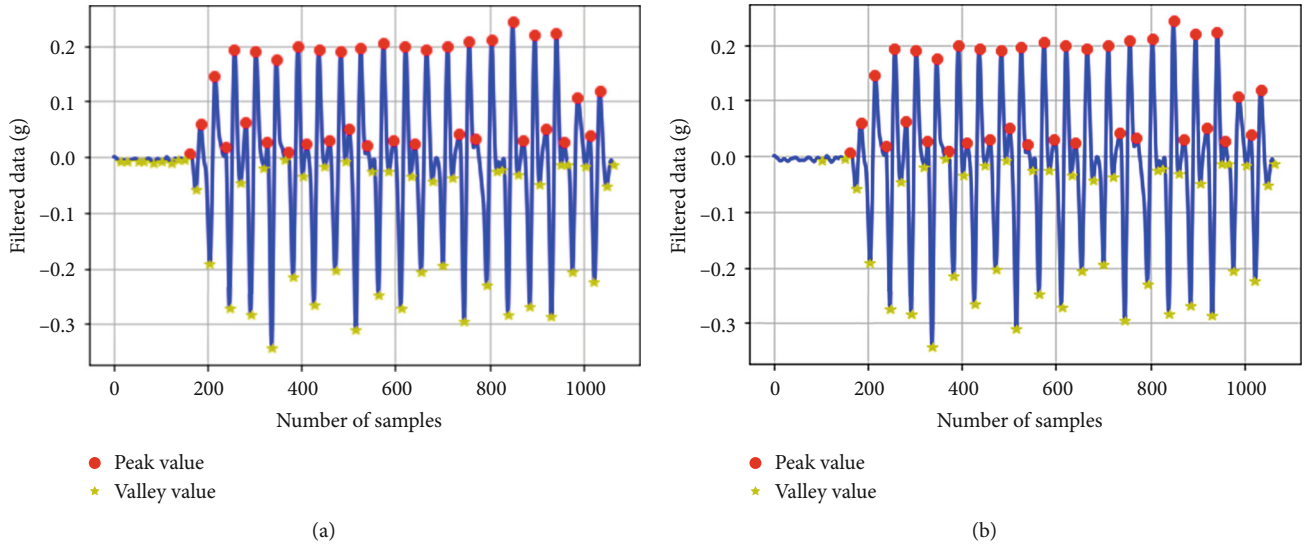


FIGURE 4: Signal extraction: (a) peak and valley identification; (b) walking signal extraction.

Similarly, for $a_valley(j)$, it will be reserved as long as the $a_valley(j)$ is satisfied as the conditions:

$$a_valley(j) \leq c * \frac{\sum_j^{j+9} a_valley(j)}{10} \&\& a_valley(j) \leq Th2. \quad (4)$$

In our proposed method, we set $Th1 = 0.02$, $Th2 = -0.02$, and $c = 0.4$, based on experience. For the last 9 peaks and valleys, the dynamic threshold $\sum_i^{i+9} a_peak(i)/10$ in formula (3) and $\sum_j^{j+9} a_valley(j)/10$ in formula (4) cannot be performed anymore, so they were replaced with the last 10th calculation result.

(ii) Minimum sampling interval

Sampling interval means the sampling number between the current peak (valley) and the next one. The pedestrian

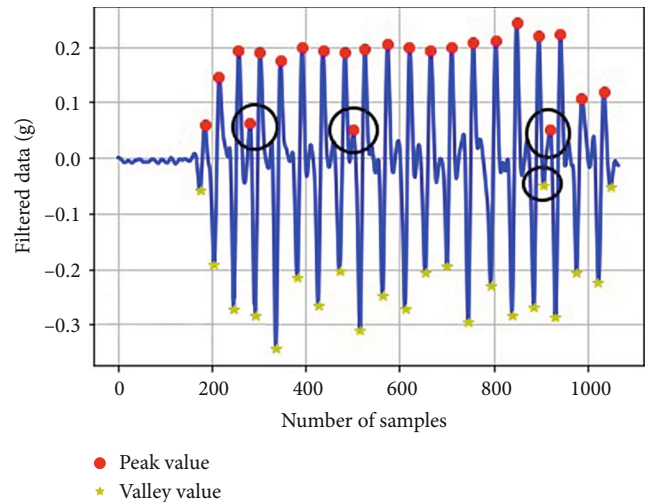


FIGURE 5: Neighboring difference outlier definition.

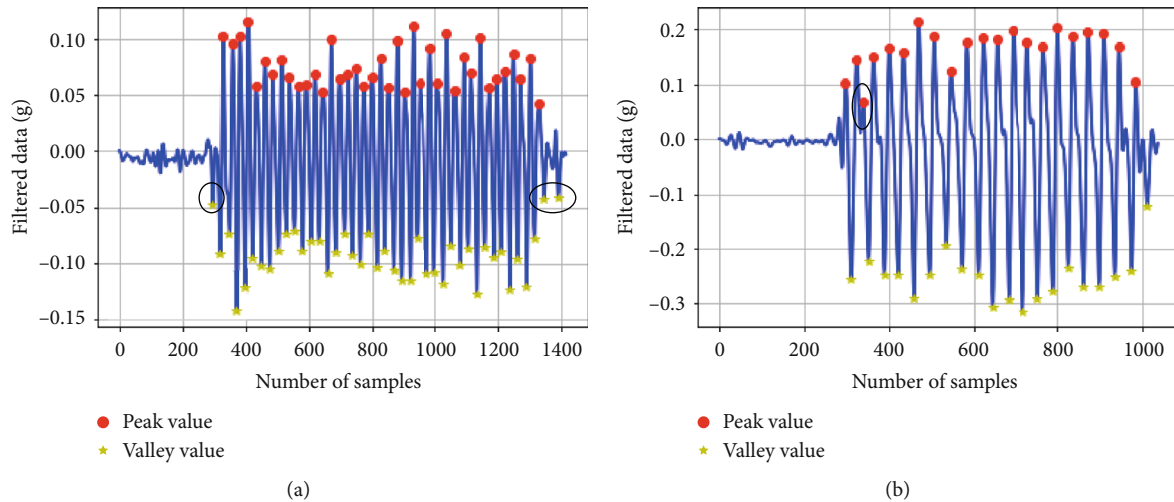


FIGURE 6: Peak-valley sequence outlier definition: (a) unpaired outliers at both ends; (b) unpaired outliers in the middle.

stride frequency ranges from 1 Hz to 3 Hz according to the data collected from 10 volunteers walking at different speeds, while the sampling rate of the device is set to 50 Hz; then, the corresponding number of sampling for each step is between 17 and 50. Therefore, if the sampling interval is less than 16, the current peak (valley) is considered to be a suspected false peak (valley), which is then compared with the next peak (valley), and the one closer to zero is deleted.

(iii) Neighboring difference

When people take large stride with low stride frequency, we detect abnormal values like the one that circled in Figure 5, which always appear between two truth values. Neighboring difference refers to the following formula:

$$\begin{cases} \Delta y_1 = y_p - y_0, \\ \Delta y_2 = y_n - y_0, \end{cases} \quad (5)$$

where Δy_1 denotes the difference between the current value y_0 and the previous one y_p and Δy_2 represents the difference between the current value y_0 and the next one y_n . For peaks, if y_0 is smaller than Δy_1 and Δy_2 , while $\Delta y_1 > 0.1$ and $\Delta y_2 > 0.1$, the current peak value is considered to be abnormal. Analogously, for valleys, if y_0 is greater than Δy_1 and Δy_2 , while $\Delta y_1 < -0.15$ and $\Delta y_2 < -0.15$, then the current valley value is regarded as the false one. Threshold values 0.1 and -0.15 are determined by the experiments in this paper.

(iv) Peak-valley sequence

Peak-valley sequence means that the peak and valley corresponding to every step should occur in pair along sampling time. Eventually, we need to eliminate two sorts of unpaired outliers: two ends of data are not paired, as shown in Figure 6(a), so the redundant values need to be removed. Another case is that the middle part of data is not paired, as shown in Figure 6(b); hence, values closer to the zero line are removed.

After eliminating the abnormal peaks and valleys, we can acquire the number of steps that is equal to the number of detected peak-valley pairs.

3. A Coefficient Self-Determined Step Length Estimator

The step length is different from person to person, and each step of an individual also varies from moment to moment. In the majority of parameter-dependent SLE models, acceleration is the most common feature to use, among which the popular Weinberg model [22] is defined as

$$\text{Step_Length} = k \cdot \sqrt[4]{a_{\max} - a_{\min}}, \quad (6)$$

where k is a constant, a_{\max} and a_{\min} are the maximum and minimum vertical acceleration in a single stride, respectively.

Kim et al. [23] proposed a method based on the average of the acceleration in each step during walking:

$$\text{Step_Length} = k \cdot \sqrt[3]{\frac{\sum_{i=1}^N |a_i|}{N}}, \quad (7)$$

where k is the user parameter, N is the number of sampling points in a step, and a_i is the vertical acceleration of i^{th} points.

Scarlet [24] developed an empirical model:

$$\text{Step_Length} = k \cdot \frac{\left(\frac{\sum_{i=1}^N |a_i|}{N}\right) - a_{\min}}{a_{\max} - a_{\min}}. \quad (8)$$

The explanations for parameters are the same as above.

However, the parameter in the step length formula above is a calibrated constant and needs to be predetermined for an individual to use, which is inconvenient and results in poor accuracy. Several studies have proposed to automatically correct the k -value in formula (6). Ho et al. [28] derived a k -factor as a polynomial function of the average step velocity.

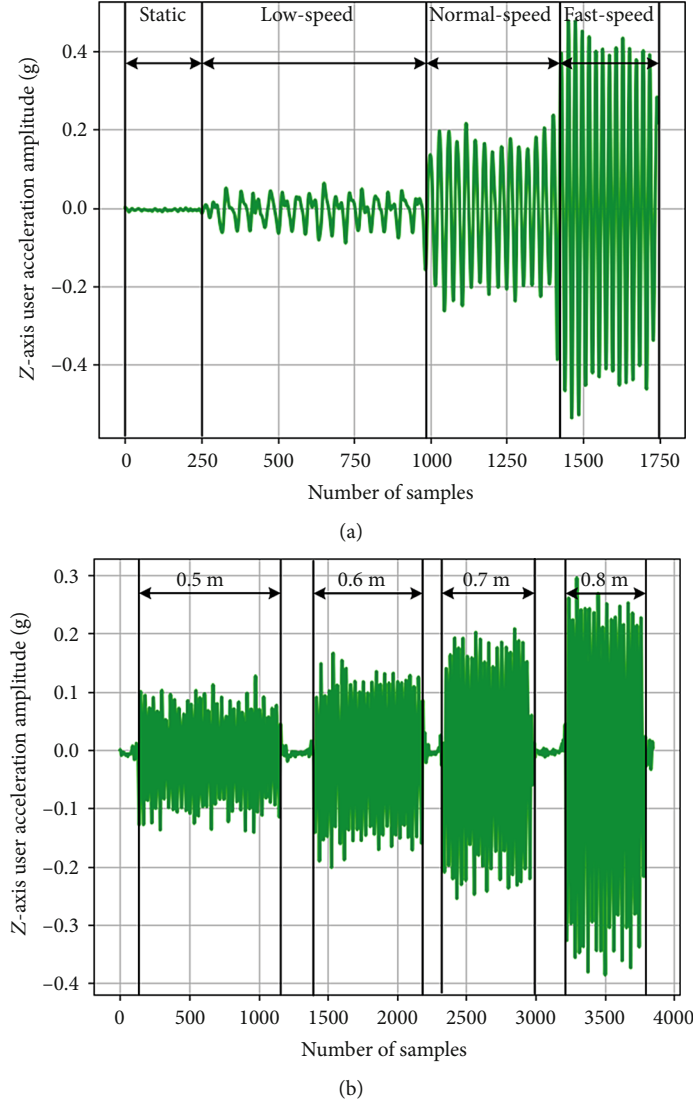


FIGURE 7: Relationship between z -axis user acceleration amplitude and stride frequency and step length: (a) acceleration amplitude at various stride frequency; (b) acceleration amplitude under different step lengths.

The adaptive k -value was obtained as follows:

$$k = 0.68 - 0.37 \times \bar{v}_{\text{step}} + 0.15 \times \bar{v}_{\text{step}}^2, \quad (9)$$

where \bar{v}_{step} was computed as the magnitude of the average velocities on three dimensional axes, x , y , and z in each step. The speeds in the three axes are obtained using acceleration double integral. Therefore, an additional filtering process is required to reduce the cumulative error caused by double integral. Strozzi et al. [29] developed the following formula to calibrate the Weinberg step length estimator:

$$k = \beta \times \frac{1}{\sqrt[3]{a_{\text{max}}(i)}}, \quad (10)$$

where β is a per-subject estimated constant and $a_{\text{max}}(i)$ is the maximum acceleration magnitude during the consid-

TABLE 1: Different combinations of step length (S_L) and stride frequency (f) and corresponding step numbers.

f (Hz)	Step number						
	1.00	1.25	1.5	1.75	2.00	2.25	2.50
S_L (m)							
0.40	40	40	40	40	40	40	40
0.47	34	34	34	34	34	34	34
0.54	29	29	29	29	29	29	29
0.61	26	26	26	26	26	26	26
0.68	23	23	23	23	23	23	23
0.75	21	21	21	21	21	21	21
0.82	19	19	19	19	19	19	19

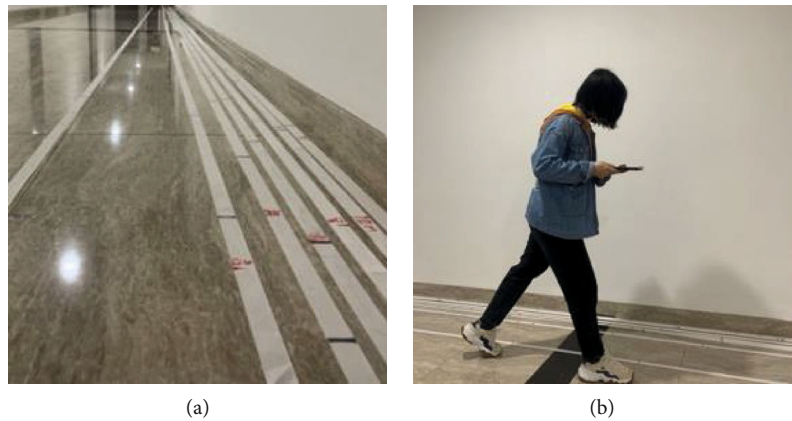


FIGURE 8: Method of walking with fixed step length: (a) the white cloth strip with black and red marks; (b) the volunteer steps on the black marks during walking.

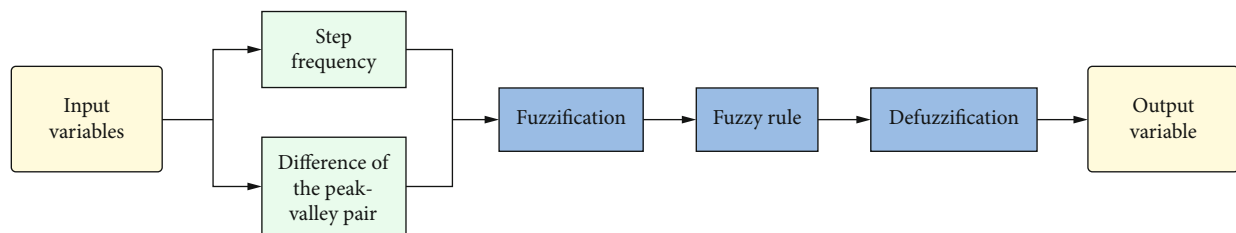


FIGURE 9: Fuzzy logic block diagram.

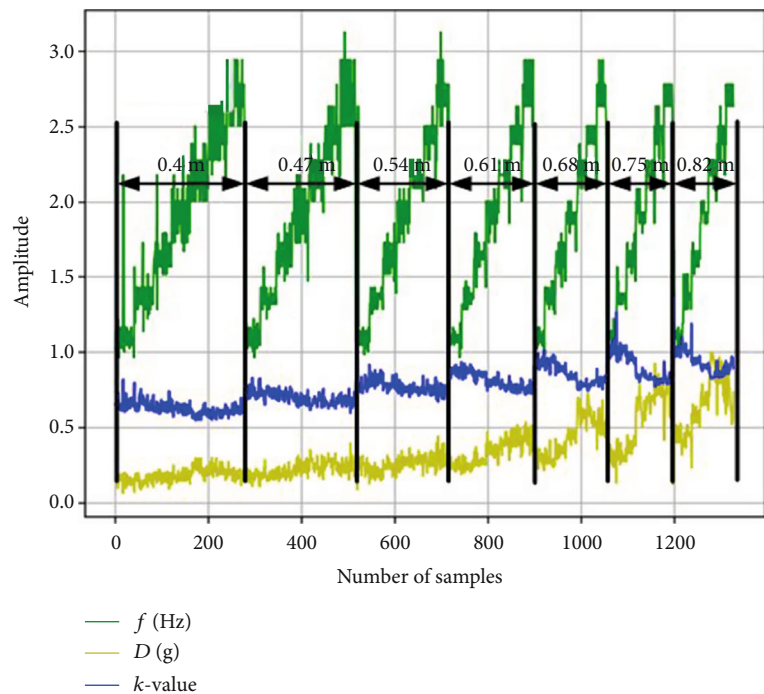


FIGURE 10: The experimental data f , D , and k of one subject.

ered i^{th} step. Although this method dynamically estimates the value of k when the pedestrian is walking, the constant β still needs to be precalibrated for different users. Therefore,

we design a fuzzy controller to adjust the k -value in Equation (6) adaptively for different kinds of pedestrians walking at various speeds, only using a simple low-pass filter.

Under static, low-speed, normal-speed, and fast-speed walking, the collected data change of the z -axis user acceleration amplitude is shown in Figure 7(a), after the low-pass filter. It can be seen from Figure 7(a) that the waveform changes from sparse to dense and the acceleration amplitude changes from small to large. From this fact, we can conclude that as the stride frequency increases, walking fluctuation in the vertical axis will become more obvious, and the amplitude reflected in the z -axis will be larger.

Besides, when walking in steps of 0.5 m, 0.6 m, 0.7 m, and 0.8 m at a stride frequency of 1.5 Hz, the amplitude change in the z -axis user acceleration is expressed in Figure 7(b). To make the step length coincide with a certain length, a white cloth strip is laid on the floor and all foot stepping points are marked with black on the strip. To adjust the stride frequency to 1.5 Hz, a metronome app is utilized. As can be seen from Figure 7(b), when the step length increases, the acceleration amplitude of the z -axis becomes larger.

According to the relationship described above, we develop a more detailed experiment to establish our fuzzy rule base. Under the fixed step length, stride frequency, and the number of steps, 10 healthy volunteers walk about 16 m holding the iPad to collect user acceleration data. The step length is set from 0.4 m to 0.82 m, every 0.07 m a group, and 7 groups in all, while the stride frequency is set from 1 Hz to 2.5 Hz, every 0.25 Hz a group, and 7 groups in all. Table 1 clarifies the number of steps that need to be performed under the combination of different step lengths and stride frequency, and each volunteer has to walk 49 times. As shown in Figure 8, for the control of fixed step length, we use a wide white cloth strip with black marks, and the volunteers stepped on the marks when walking. The red labels are marked on the cloth strips, which are aligned with the ground stamps, to avoid the error caused by each laying.

Figure 9 shows the fuzzy logic block diagram in this study. We determine the difference of the peak-valley pair obtained in Section 2 and stride frequency as input variables and the k -value in Equation (6) as an output variable. The final output k is brought into Equation (6) to estimate step length for each step detected. The following is the detail of FL:

(i) Input variables and output variable

The input variables of FL are the stride frequency f and the difference D of the peak-valley pair:

$$\begin{cases} f_n = \frac{1}{T_n}, \\ D_n = a_{\text{peak}(n)} - a_{\text{valley}(n)}, \end{cases} \quad (11)$$

where T is the time required for the pedestrian walking in a step and $a_{\text{peak}(n)}$ and $a_{\text{valley}(n)}$ are the peak and valley in the n^{th} step obtained by peak-valley detection.

TABLE 2: Fuzzy control rule base.

f D	k						
	NB	NM	NS	ZE	PS	PM	PB
NB	NS	NS	NS	NM	NB	NB	NB
NM	PS	PS	PS	ZE	ZE	NS	NM
NS	PM	PM	PS	PS	ZE	ZE	NS
ZE	PB	PM	PS	PS	PS	ZE	ZE
PS	PB	PM	PS	PS	PS	ZE	ZE
PM	PB	PM	PS	PS	PS	PS	PS
PB	PB	PM	PM	PS	PS	PS	PS

The domain of f is determined to be $[1, 2.8]$, while D is determined to be $[0.12, 0.9]$ based on data collected by each volunteer walking 49 times. The output variable k -value is calculated by the following formula:

$$k = \frac{S_L}{\sqrt[4]{a_{\text{max}} - a_{\text{min}}}}, \quad (12)$$

where S_L is the step length that volunteers perform according to Table 1. Based on the walking data of 10 volunteers, the calculation result shows that the domain of output variable k is $[0.55, 1.0]$.

As the quantization level increases, the output accuracy will be improved, but the complexity of the calculation will also increase. Taking into account the complexity of the calculation and the accuracy of the output variables, the quantization levels of the input and output variables in this paper are all seven levels relative to experimental setup above, i.e., $f, D, k = \{-3, -2, -1, 0, 1, 2, 3\}$, as well as the fuzzy subsets adopted seven levels: NB, NM, NS, ZE, PS, PM, and PB (N: negative, B: big, M: middle, S: small, P: positive, ZE: zero). For membership function, we employ triangular function.

(ii) Rule base

In this paper, the adopted fuzzy rules are Mamdani-based. The experimental data of one subject performing 49 times as shown in Figure 10, from which k is changeable and significantly relevant to f and D . At the same step length, as f increases, D becomes larger and k is smaller instead. Moreover, at the same stride frequency f , as step length increases, both D and k get larger. We establish the fuzzy control rule base according to the commonality of these data changes, as shown in Table 2.

(iii) Defuzzification

The center-of-gravity method is used as a defuzzification strategy in this paper, which has a smoother output inference control; i.e., corresponding to a small change in the input signal, the final output of its inference generally also changes to some extent [31].

TABLE 3: The error of the peak-valley detection and peak detection on 500 steps of free walking.

	Volunteer	Gender	True steps	Peak-valley detection		Peak detection	
				Estimated	Error	Estimated	Error
Two of the ten volunteers	Vol 1	Male	500	500	0	507	1.4%
	Vol 2	Female	500	499	0.2%	509	1.8%
Additional five volunteers	Tester 1	Male	500	500	0	507	1.4%
	Tester 2	Male	500	500	0	505	1.0%
	Tester 3	Male	500	501	0.2%	510	2.0%
	Tester 4	Male	500	498	0.4%	506	1.2%
	Tester 5	Female	500	496	0.8%	505	1.0%

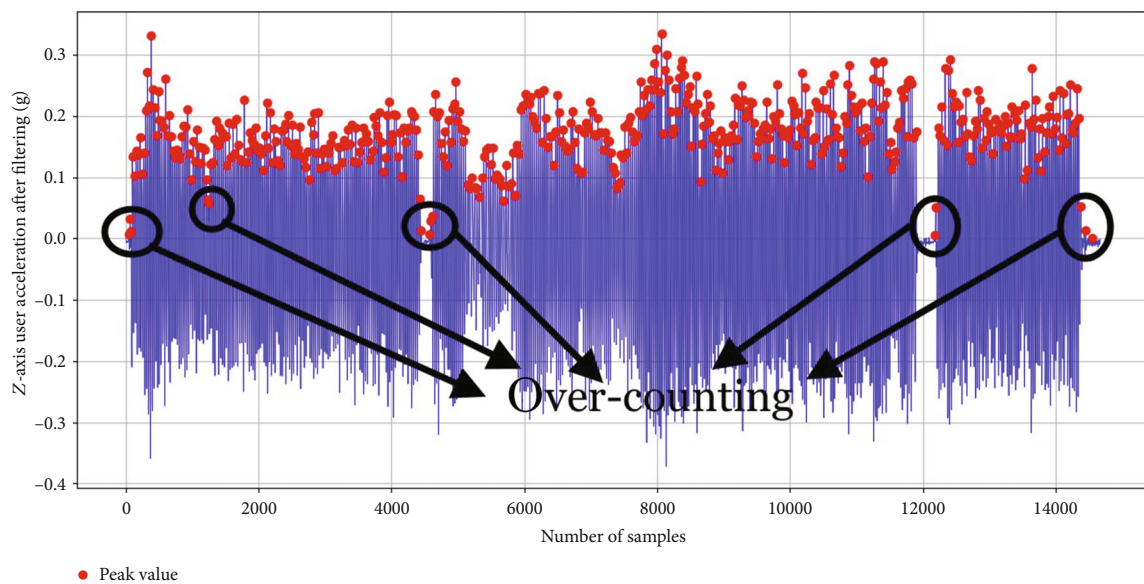
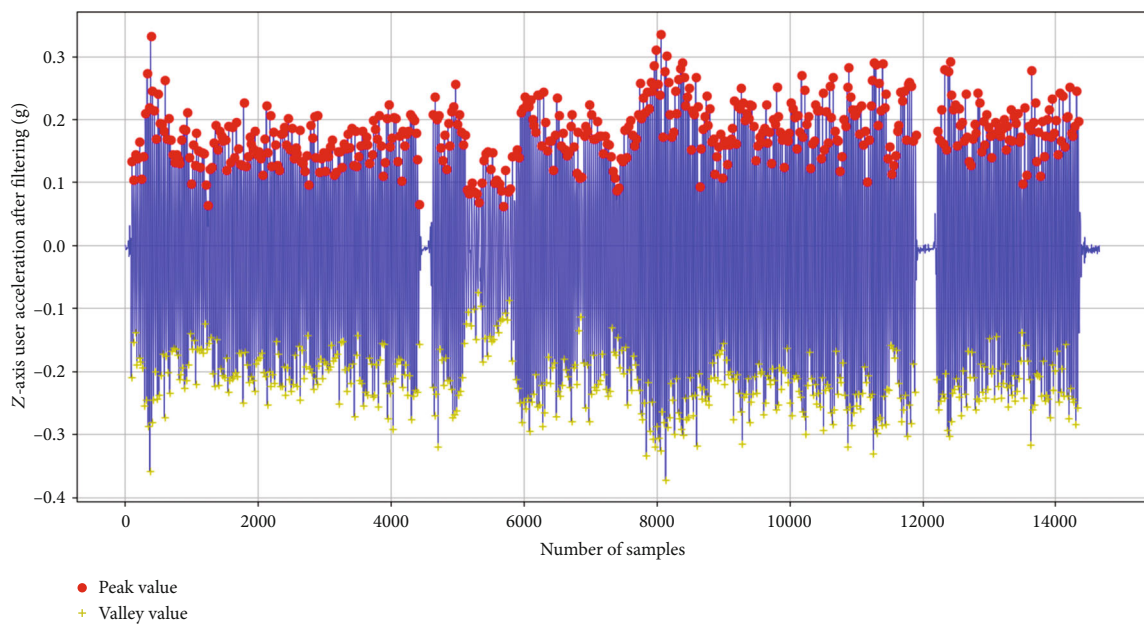


FIGURE 11: Results after executing 500 steps in free walking: (a) step counting results based on the peak-valley detection; (b) step counting results based on peak detection.

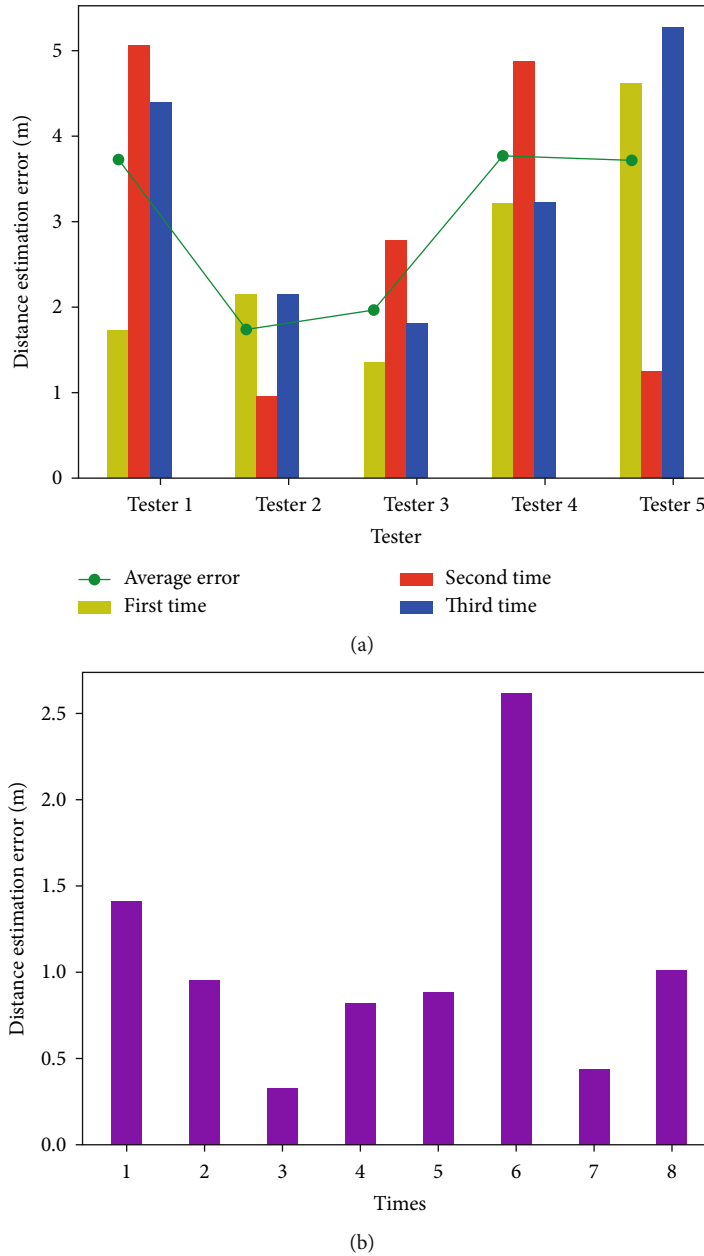


FIGURE 12: Distance estimation results: (a) distance estimation error of 5 testers walking freely 100 m; (b) distance estimation error of walking 60 m when holding the iPad at an angle about 20 degrees.

4. Experiments and Results

4.1. Evaluation of the Peak-Valley Detection. To evaluate the performance of the peak-valley detection, we have used the following formula to estimate the error of the proposed method:

$$\text{Error} = \frac{|E - T|}{T} \times 100\%, \quad (13)$$

where E is the estimated number of steps by the peak-valley detection and T is the true steps. We randomly select 2 volunteers in the experiment above and 5 additional volun-

teers (aged from 22 to 26, height from 1.65 m to 1.85 m, weight from 55 kg to 78 kg) to walk freely (not at a fixed stride frequency and step length) 500 steps. Table 3 shows the error of the peak-valley detection and peak detection method. It can be seen that step counting based on the peak detection is subject to overcounting problem. By calculating the mean error of seven people, the average accuracy of the peak-valley detection method is as high as 99.77%. Compared with the traditional peak detection method, the average accuracy is improved by 1.17%.

Figure 11 shows the results after executing 500 steps in free walking using the peak-valley detection and peak detection. 498/500 steps were detected based on the peak-valley

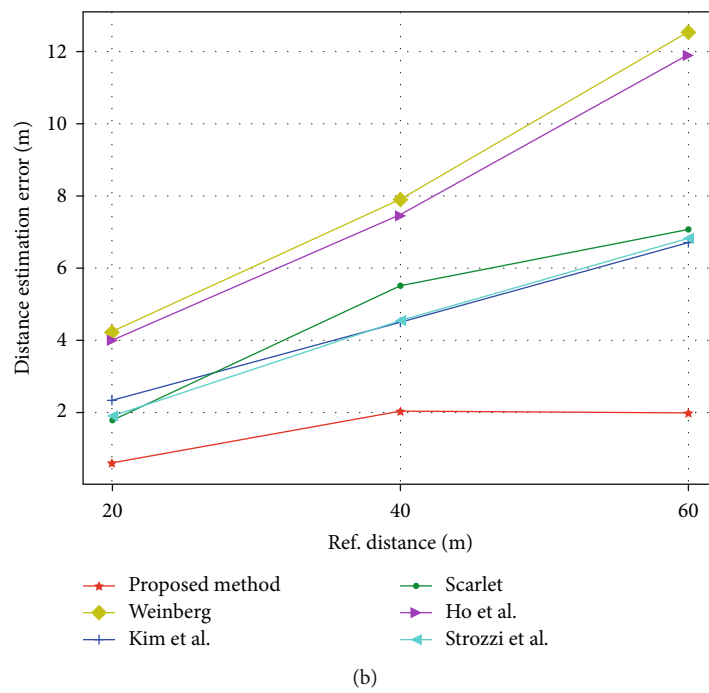
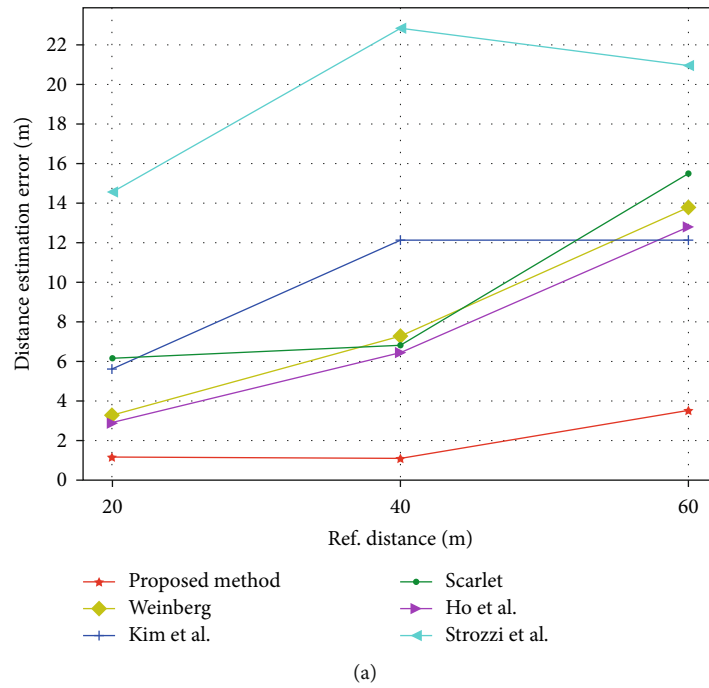


FIGURE 13: Continued.

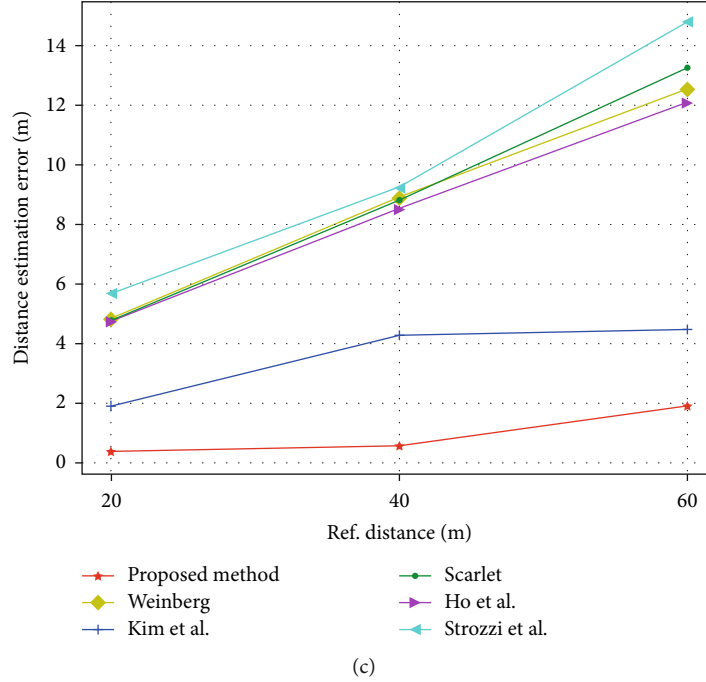


FIGURE 13: Distance estimation error at three different walking speeds of six step length estimators: (a) low speed; (b) normal speed; (c) high speed.

detection method, and 515/500 steps based on peak detection. Comparing Figures 11(a) and 11(b), we can figure out that the emergence of the overcounting problem based on peak detection mainly occurs at the start walking phase and stop walking. Our proposed method can effectively overcome this issue.

4.2. Evaluation of Improved Step Length Estimation Model. To test the performance of our improved SLE model, we ask 5 testers to walk freely three times in a corridor and every time stop at approximately 100 m. The estimated distance is obtained by accumulating each step length. The following formula is used to calculate the error of the estimated distance:

$$\text{Error} = |D_E - D_T|, \quad (14)$$

where D_E is estimated distance and D_T is the true walking distance.

As shown in Figure 12(a), among the 5 testers, the error of tester 2 is only 0.95 m in the walk of the second time, while the error of tester 5 in the walk of the third time is the largest, which is 5.27 m. Moreover, the average error is the mean error of each tester walking three times, from which we can see that the average error of 5 testers is all less than 4 m and the smallest one is 1.74 m. Therefore, the proposed method is adaptive for different people with free walking.

In daily walking, pedestrians cannot hold the device fully horizontally while navigating. We also test the accuracy of walking distance estimation when holding an iPad at an angle (about 20 degrees). Figure 12(b) indicates that the average error of a tester walking 8 times is 1.06 m per 60 m when

tilting the iPad. Consequently, it is effective for pedestrians when holding the devices at an angle to estimate the walking distance based on the method proposed in this paper.

Figure 13 shows the distance estimation error (m) for a tester conducted walking experiments covering three distances (20 m, 40 m, and 60 m) at three different walking speeds (low, normal, and high) of six step length estimators. The results indicate that our proposed method can obtain better accuracy of distance estimation for various walking speeds. Particularly, as the walking distance rises, the distance estimation error of our proposed method increases slightly. However, the other five SLE models increase significantly; especially when walking 60 m at low speed, the distance estimation error is over 12 m. Ho et al.'s [28] method has a certain improvement on the Weinberg's method [22], but the effect is weak. Strozzi et al.'s [29] method acquires the highest error when walking at low speed and high speed, which means that the parameter β in formula (10) still needs to be dynamically adjusted at different walking speeds. Therefore, our proposed method can significantly reduce the estimation error and is suitable for the pedestrian walking at various speeds and long distance.

5. Discussion

Based on the proposed method, we conduct experiments on different walking surfaces (slope, downstairs, upstairs) and different movement modes (running, jogging, lateral walking, backward walking). Table 4 shows the results of walking on different surfaces. The experiment of the slope surface is carried out on a treadmill with the inclination angle of 9 degrees. The width of the stair is 27 cm and the true distance

TABLE 4: The results of walking on different surfaces.

Surface type	True steps	Detected steps	Step detection error (%)	True distance (m)	Estimated distance (m)	Distance estimation error (m)
Slope	138	138	0	60	70.25	10.25
Downstairs	108	105	2.78	30.95	69.73	38.78
Upstairs	101	96	4.95	30.95	71.41	40.46

TABLE 5: The results of walking at different movement modes.

Movement mode	True steps	Detected steps	Step detection error (%)	True distance (m)	Estimated distance (m)	Distance estimation error (m)
Running	31	11	64.52	30	10.23	19.77
Jogging	35	21	40.00	30	16.40	13.60
Lateral walking	45	44	2.22	30	29.32	0.68
Backward walking	63	63	0	30	33.85	3.85

of stair experiment is the sum of all walked stairs and intermediate rest platform. From Table 4, we can conclude that our peak-valley detection method is still applicable, but the fuzzy rule base needs to be adjusted when walking on the slope or stairs. The case of undercounting occurs when the pedestrian is walking on the intermediate rest platform which is the plane, because the parameter c in the stage of dynamic threshold is ineffective in this situation. Table 5 expresses the results of walking at different movement modes, from which we can know that our proposed method still applies to lateral walking and backward walking. However, there is a significant error when the pedestrian is running or jogging. The stride frequency of running or jogging exceeds 3 Hz, while the proposed method only considers the walking frequency ranging from 1 Hz to 3 Hz. Therefore, the stage of minimum sampling interval leads to the severe problem of undercounting.

In the future work, to expand the application scenarios of our method, we will classify and recognize pedestrian movement patterns (running, walking, upstairs, downstairs, slope, etc.) and self-adjust the fuzzy rule base to the corresponding patterns.

6. Conclusions

This paper proposes a coefficient self-determined step length estimation method based on the peak-valley detection. Employing the peak-valley detection, the abnormal values will be filtered out effectively. Especially, it is significantly effective for the overcounting problem in the peak detection. Based on the FL, a fuzzy controller is established, which can adjust the k -value in Weinberg's model adaptively for every step detected by the peak-valley detection. The experimental results confirm that our proposed method is universal for different types of pedestrians. Compared with three typical nonlinear models and two state-of-the-art methods, the proposed method has higher accuracy when walking at various step velocities and long distance.

Data Availability

The inertial data of the iPad used to support the results of this study can be obtained from the authors of this article (e-mail: wenzialu@sues.edu.cn (Wenzia Lu); njuptzhuh@163.com (Hai Zhu)).

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the team of Fusion Computing and Location Services. This research was funded by the Key Project of Science and Technology Commission of Shanghai Municipality (Grant No. 18511101600), the "Sailing Program" of Young Science and Technology Talents supported by Shanghai Science and Technology Commission (Grant No. 19YF1418200), the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 61902237), Natural Science Foundation of Shanghai (Grant no. 17ZR1411900), and the Opening Project of Shanghai Key Laboratory of Integrated Administration Technologies for Information Security (Grant no. AGK2015006).

References

- [1] S. Khruahong, X. Kong, K. Sandrasegaran et al., "Multi-level indoor navigation ontology for high assurance location-based services," in *2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE)*, pp. 128–131, Singapore, Singapore, January 2017, IEEE Xplore.
- [2] L. Chen and H. Hu, "IMU/GPS based pedestrian localization," in *2012 4th Computer Science and Electronic Engineering Conference (CEECE)*, pp. 23–28, Colchester, UK, September 2012, IEEE Xplore.

- [3] K. Mao and G. Chen, "The application status quo and development trend of indoor positioning technology," *Modern Surveying and Mapping*, vol. 41, no. 5, pp. 31–34, 2018.
- [4] X. Li, J. Wang, and C. Liu, "A Bluetooth/PDR integration algorithm for an indoor positioning system," *Sensors*, vol. 15, no. 10, pp. 24862–24885, 2015.
- [5] Q. Fan, B. Sun, Y. Sun, and X. Zhuang, "Performance enhancement of MEMS-based INS/UWB integration for indoor navigation applications," *IEEE Sensors Journal*, vol. 17, no. 10, pp. 3116–3130, 2017.
- [6] W. Chai, C. Chen, E. Edwan, J. Zhang, and O. Loffeld, "INS/Wi-Fi based indoor navigation using adaptive Kalman filtering and vehicle constraints," in *2012 9th Workshop on Positioning, Navigation and Communication*, pp. 36–41, Dresden, Germany, March 2012, IEEE Xplore.
- [7] R. Li, J. Zhang, D. Xu, X. X. Chen, and Q. L. Xu, "Micro-electro-mechanical system-inertial measurement unit indoor pedestrian dead reckoning based on motion classification and step frequency adjustment," *Journal of Shanghai University*, vol. 20, no. 5, pp. 612–623, 2014.
- [8] A. R. Jimenez, F. Seco, C. Prieto, and J. Guevara, "A comparison of pedestrian dead-reckoning algorithms using a low-cost MEMS IMU," in *2009 IEEE International Symposium on Intelligent Signal Processing*, pp. 37–42, Budapest, Hungary, August 2009, IEEE Xplore.
- [9] G. Panahandeh, N. Mohammadiha, A. Leijon, and P. Handel, "Continuous hidden Markov model for pedestrian activity classification and gait analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 5, pp. 1073–1083, 2013.
- [10] E. M. Diaz, A. L. M. Gonzalez, and F. de Ponte Müller, "Standalone inertial pocket navigation system," in *2014 IEEE/ION Position, Location and Navigation Symposium - PLANS 2014*, pp. 241–251, Monterey, CA, USA, May 2014, IEEE Xplore.
- [11] J. C. Alvarez, D. Alvarez, A. López, and R. C. González, "Pedestrian navigation based on a waist-worn inertial sensor," *Sensors*, vol. 12, no. 8, pp. 10536–10549, 2012.
- [12] Y. Wu, H. Zhu, Q. X. du, and S. M. Tang, "A survey of the research status of pedestrian dead reckoning systems based on inertial sensors," *International Journal of Automation and Computing*, vol. 16, no. 1, pp. 65–83, 2019.
- [13] K. Tumkur and S. Subbiah, "Modeling human walking for step detection and stride determination by 3-axis accelerometer readings in pedometer," in *2012 Fourth International Conference on Computational Intelligence, Modelling and Simulation*, pp. 199–204, Kuantan, Malaysia, September 2012, IEEE Computer Society.
- [14] J. Seo and T. Laine, "Accurate position and orientation independent step counting algorithm for smartphones," *Journal of Ambient Intelligence and Smart Environments*, vol. 10, no. 6, pp. 481–495, 2018.
- [15] V. T. Pham, D. A. Nguyen, N. D. Dang et al., "Highly accurate step counting at various walking states using low-cost inertial measurement unit support indoor positioning system," *Sensors*, vol. 18, no. 10, p. 3186, 2018.
- [16] X. Kang, B. Huang, and G. Qi, "A novel walking detection and step counting algorithm using unconstrained smartphones," *Sensors*, vol. 18, no. 1, p. 297, 2018.
- [17] F. Gu, K. Khoshelham, J. Shang, F. Yu, and Z. Wei, "Robust and accurate smartphone-based step counting for indoor localization," *IEEE Sensors Journal*, vol. 17, no. 11, pp. 3453–3460, 2017.
- [18] B. Ao, Y. Wang, H. Liu, D. Li, L. Song, and J. Li, "Context impacts in accelerometer-based walk detection and step counting," *Sensors*, vol. 18, no. 11, p. 3604, 2018.
- [19] W. Shao, H. Luo, F. Zhao, C. Wang, A. Crivello, and M. Z. Tunio, "DePedo: Anti periodic negative-step movement pedometer with deep convolutional neural networks," in *2018 IEEE International Conference on Communications (ICC)*, pp. 1–6, Kansas City, MO, USA, May 2018, IEEE.
- [20] A. R. Pratama, Widyawan, and R. Hidayat, "Smartphone-based pedestrian dead reckoning as an indoor positioning system," in *2012 International Conference on System Engineering and Technology (ICSET)*, pp. 1–6, Bandung, Indonesia, September 2012, IEEE Xplore.
- [21] V. Renaudin, M. Susi, and G. Lachapelle, "Step length estimation using handheld inertial sensors," *Sensors*, vol. 12, no. 7, pp. 8507–8525, 2012.
- [22] H. Weinberg, *Using the ADXL202 in Pedometer and Personal Navigation Applications*, Analog Devices, Inc., Norwood, MA, USA, 2002.
- [23] J. W. Kim, H. J. Jang, D. H. Hwang, and C. Park, "A step, stride and heading determination for the pedestrian navigation system," *Journal of Global Positioning System*, vol. 3, no. 1&2, pp. 273–279, 2004.
- [24] J. Scarlet, *Enhancing the Performance of Pedometers Using a Single Accelerometer*, Analog Devices, Inc., Norwood, MA, USA, 2005.
- [25] H. Xing, J. Li, B. Hou, Y. Zhang, and M. Guo, "Pedestrian stride length estimation from IMU measurements and ANN based algorithm," *Journal of Sensors*, vol. 2017, Article ID 6091261, 10 pages, 2017.
- [26] J. Hannink, T. Kautz, C. F. Pasluosta et al., "Mobile stride length estimation with deep convolutional neural networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 2, pp. 354–362, 2018.
- [27] Y. C. Lai, C. C. Chang, C. M. Tsai, S. C. Huang, and K. W. Chiang, "A knowledge-based step length estimation method based on fuzzy logic and multi-sensor fusion algorithms for a pedestrian dead reckoning system," *International Journal of Geo-Information*, vol. 5, no. 5, p. 70, 2016.
- [28] N.-H. Ho, P. Truong, and G.-M. Jeong, "Step-detection and adaptive step-length estimation for pedestrian dead-reckoning at various walking speeds using a smartphone," *Sensors*, vol. 16, no. 9, p. 1423, 2016.
- [29] N. Strozzi, F. Parisi, and G. Ferrari, "A novel step detection and step length estimation algorithm for hand-held smartphones," in *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–7, Nantes, France, September 2018, IEEE.
- [30] <https://developer.apple.com/documentation/coremotion>.
- [31] W. Wei, *Intelligent control technology*, Beijing, China, 2015.

Research Article

Design of a Secure Wireless Home Automation System with an Open Home Automation Bus (OpenHAB 2) Framework

Robert A. Sowah , **Dale E. Boahene**, **Dalton C. Owoh**, **Rexford Addo**, **Godfrey A. Mills**, **Wiafe Owusu-Banahene**, **Gifty Buah**, and **Baffour Sarkodie-Mensah**

Department of Computer Engineering, University of Ghana, Accra, Ghana P.O. Box LG 77, Legon

Correspondence should be addressed to Robert A. Sowah; rasowah@ug.edu.gh

Received 28 June 2020; Revised 4 August 2020; Accepted 11 September 2020; Published 30 October 2020

Academic Editor: Rafael Morales

Copyright © 2020 Robert A. Sowah 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.

There is rapid interest growing in the use of smart, connected devices. The developing world market for smart technology is evolving to adopt and adapt to the interconnected world of devices leading to the Internet of Things (IoT) everywhere. This research paper presents the design, development, and deployment of a prototype for the secure wireless home automation system with OpenHAB 2. We employed the use of two (2) high-performance microcontrollers, namely, the Arduino Mega 2560, interfaced with a 16-channel relay, and Raspberry Pi Model B, running the OpenHAB software. The Raspberry Pi functioned as the server to develop a prototype of an automated smart home that is remotely controllable from both a web application and an Android mobile app. In designing a wireless controlled switch for home appliances, two security procedures were implemented, namely, the token-based JSON Web Token (JWT) interface and Advanced Encryption Standard (AES) procedures for authentication and data encryption. Our system delivered a home automation system that leverages on the power of the latest version of OpenHAB to maximize productivity and overall home security while making it adaptable to the management of individual devices. When tested, both the developed hardware and software modules performed extremely well to meet the goal of a secured home automation system. Industry-standard penetration testing tools and frameworks, including Aircrack-ng, were utilized; wireless network audit began with a full sweep of the wireless frequencies with excellent results. It also ensures the efficient use of energy in the home as devices are intelligently controlled from both mobile and web applications. The results of the design and implementation of the additional layer for the security of the OpenHAB framework provide various theoretical and practical implications for home automation.

1. Introduction

Home automation is the adoption of a system to control lighting, atmospheric conditions, entertainment systems, surveillance systems, and home appliances. It allows for devices in the home to be connected to a remotely controllable network. This technology makes life easier for the user and saves energy by utilizing devices with the most priority and importance [1–5]. Controls can be made as easy as turning off lights with a remote or as complex as setting up a network of items [6] that can program a controller via smartphones from everywhere in the world. It allows for devices in the home to be connected to a remotely controllable network.

The idea of home automation has been around for decades [7], but only in recent times were actual architectural designs being implemented. The earliest form of home automation dates back to 1893 when the first television with a remote control system was patented. The popularity of home automation grew afterward in the early 2000s due to an increase in the demand for various technologies. The need for home automation has continued to increase as more affordable domestic technologies emerge in the market today.

Developing one's smart devices to work with all home automation systems would prove a difficult task as it would require the assistance of the smartphone companies to properly set up these devices to communicate with the home automation systems. This research paper uses OpenHAB 2—the

latest stable version of the software with the Raspberry Pi 3 Model B and Arduino Mega microcontroller to deliver a prototype of a secure wireless home automation system.

In recent times, there have been attempts at bridging the gap between compatibility with appliances of different vendors. A quick but rather inefficient fix for this problem would be to have separate applications for each kind of device [8–10]. OpenHAB serves to overcome most of the common integration issues associated with wireless home automation. It integrates several cross-platform home automation systems into a single solution, acts as a common communicative language among various devices, makes interdevice communication neutral, and makes device integration easy [9, 10].

With recent technological developments, there has been an ongoing debate on the need to switch from traditional home settings to secure automated homes. Most homeowners in developing countries would see this new initiative as an unwarranted luxury. They would point to the cost of installations and the lack of an enabling environment as a significant hindrance. It should, however, be noted that improvements in home automation systems would improve productivity and security in households at affordable costs.

The drawbacks of the traditional home systems are highlighted below: (1) the repetitive nature of most tasks creates room for inefficiency and time-wastage and thus makes the use of home automation essential; (2) the regular traditional homes cannot be easily monitored or controlled. Events going on at different parts of the home cannot be monitored and controlled effectively. The status of the home appliances cannot be checked without the homeowner being physically present; and (3) the cost of installing and maintaining security devices in traditional homes is high, and therefore, inexpensive options to home security should be harnessed [2, 10, 11].

Most homeowners in developing countries lack complete and total control over their homes. They are not able to access vital home automation features such as control and monitoring of home appliances, low-cost security, and efficient energy usage by implication. This paper seeks to explore a way to replace traditional home systems with secure wireless automated systems.

The OpenHAB 2 protocol is relevant to automating the home while interconnecting appliances based on its flexibility and capability for full customization. Therefore, the following objectives are pursued within the context of this research work, namely, (1) provision of an easy control mechanism for home appliances via a mobile or web application; (2) improvement of the security of connected home appliances through the system's inbuilt intrusion detection, alarm, and wireless communication data encryption standards; (3) meeting the essential energy management requirement of the household by providing a means to monitor and remotely turn off unnecessary active appliances to conserve energy and reduce electricity bills; (4) offering a wirelessly controlled switch for all home appliances; and (5) making a significant contribution to the existing body of knowledge on secured home automation.

There are key factors that drive sustainable automation using OpenHAB software. These are specified below:

- (1) *Security*. The OpenHAB software does not enforce any access control mechanism for its users and depends solely on the security of the wireless network. It can, however, be secured through user authentication and authorization.
- (2) *Cost-Effectiveness*. The initial cost of installation is relatively higher than in conventional homes. It has a high maintenance cost. However, inbuilt energy-saving mechanisms could save some running costs and enables optimal utilization of electrical utilities.
- (3) *Alignment to Existing Environmental Infrastructure*. It is not suitable in areas without 24-hour power supply. It becomes a high-security risk, and extra cost can be incurred if external power generating equipment is acquired.
- (4) *Flexibility in Application and Adaptation*. The OpenHAB software is open-source software that ensures flexibility in the adaptation of smart devices. It serves as a hub that brings together a diverse range of heterogeneous devices. Smart home technologies are flexible when it comes to the accommodation of new devices, appliances, and other technologies.
- (5) *OpenHAB Is Central to Smart Homes*. Like electric cars, smart grid, and other next-generation technologies, the smart home promises high market value in the near future. Presently, in most underdeveloped and developing communities, smart homes may be an imprudent acquisition as other supporting platforms are not already in place.
- (6) *Legal Factors*. In many areas, smart home technology is moving faster than the laws governing such technology. For instance, smart homes often come equipped with high-tech security features, such as audio and video recording devices. In many states and jurisdictions, prior consent is required before making an audio or video recording of someone, for instance.

Additionally, a summary of the countries who have adopted smart home automation technology is given in Table 1.

In Africa, however, smart home technology is still at a very infant stage. Nairobi in Kenya is regarded as the smartest city in Africa [13]. South Africa is already implementing smart home technologies, including other African countries like Nigeria, Ghana, Rwanda, and Ethiopia. The calendar specification of the trend in smart homes reveals an upward trend [14].

The invention of home appliances started during the first two decades of the 20th century. The middle of the 20th century saw technological advancement with the invention of the Echo IV and the kitchen computer in 1966. The Echo IV was the first smart device, although it was never

TABLE 1: Cross country summary on smart homes [12].

Rank (number of smart homes)	Country	Number of smart homes (millions)	Percentage of smart homes (% of total homes)
1	USA	40.3	32
2	China	19.3	4.9
3	Japan	7	15
4	Germany	6.1	15.7
5	UK	5.3	19.7
6	South Korea	4	20
7	India	2.2	<1
8	Australia	1.8	19.12
9	Brazil	1.2	1.9
10	Russia	0.9	1.7

commercially sold. It could compute shopping lists and turn appliances on and off. The early 1990s saw the innovation rendered by gerontechnology, which was the invention of technology to aid gerontology, e.g., medical alarms for the aged. In the 1998 and early 2000s, smart home technology took a boom for its creative innovations by different vendors. The home automation market was estimated at US\$5.77 billion in 2013 and is predicted to reach a market value of US\$12.81 billion by the year 2020 [15].

In-depth research was done on existing works and systems that had been implemented around the scope of the research work defined in this paper, namely, to design secure wireless home automation based on privacy by design. The focus of the literature review was to acknowledge the work that was done, establish their strengths and gaps, and access how different ecosystems influence outcomes. Among the many papers of interest, the under listed were considered of direct interest to this proposed research work [1–3, 7–25] and thus were reviewed. Their relative strengths and weaknesses are highlighted in the related work section.

The unique contributions of this paper are (1) design and development of cost-effective, secure home automation using the OpenHAB 2 framework with capability for device programming and customizations and (2) development of mobile and web applications for energy management and switching of connected home devices and interactive visual interface for home automation. It leverages the developed hardware and software modules to provide optimal energy management for the home. While there are studies that explored the application of OpenHAB in home automation, to the best of our knowledge, there is none that we found in the published literature that addresses the prevailing security challenges of the default OpenHAB server infrastructure, especially for connections over the Internet for smart home automation. In this paper, we implemented additional security layers of user authentication and authorization while keeping the overall cost of implementation low and maintaining the ease of deployment for everyday home use. Per our server configuration, the OpenHAB communication through the Internet is made through the JSON Web Token authentication procedure. This process makes it difficult for user identity to be hijacked by a malicious attacker. This approach proved to be more secure than the default Open-

HAB server configuration. Consequently, we leveraged on the open-source OpenHAB REST API to develop a mobile or web application that is flexible and easily adaptable for traditional home use.

Section 2 presents the literature review on smart home automation systems using different techniques and algorithms for energy management and control. It addresses the various home automation technologies with implemented web and mobile apps. It identifies the strengths and weaknesses of such systems. It proposes a novel method using the AES and JWT security protocols for enhanced security and ensures privacy by the design of home automation. The novelty of the system design architecture and methodology, including the concepts, algorithms, activity diagrams, and flowcharts, is presented in Section 3. This section highlights the various modules with their design and procedures for implementation. Section 4 focuses on the actual design implementations with circuit diagrams and simulations done in Proteus software, experimental verification, and corresponding integration of the developed modules. This section provides sufficient details on the testing and performance evaluation of the proposed secure home automation system with associated modules for efficient monitoring and control of home appliances. Finally, Section 5 provides conclusions on the system design and methodology and relevant recommendations for future enhancements.

2. Review of Related and Existing Systems

Jin et al. [1] implemented a remotely controlled home automation system based on the wireless sensor network and an embedded system with GPRS integration. The system was unique for good reasons; it allowed users to control the equipment in the home and collect data about appliances' status as well as weather conditions. Additionally, the system possessed powerful security features like the inbuilt fire detection and home intrusion detection and notification system. The system is tightly integrated with the Chinese instant mobile messaging service.

A significant drawback of this system is the absence of a mobile or web application interface to control the home appliances. The reliance of the SMS system alone for

communication meant that if the GSM device is misplaced, access control will no longer be valid.

The authors in [6] designed a system to give the homeowner complete control over their home using the Short Message Service (SMS) communicating over GSM. In this system, a GSM modem is connected to an RS232 interfaced with a MAX232 that is also connected to a PIC microcontroller, as the RS232 is not compatible with the microcontroller. When a user sends an SMS message, it is sent to the RS232, and the MAX232 converts the text messages to a TTL signal for the PIC microcontroller. The PIC controls the relay, which is the connection point for all the home appliances [6].

The gap in this system is that it is highly intrusive and requires a rewiring of electrical appliances for configuration purposes. The system is also limited in the use of only a GSM, and a SIM card is needed to control the home remotely, and if the GSM or SIM is misplaced, the system ceases to function. Additionally, no security standards were put in place to secure the system against intruders and hackers.

Kodali et al. [7] proposed the IoT-based smart security and home automation system, which utilizes certain features of the IoT domain for both communication and control of devices while ensuring their secure operation without interruption. It implemented a security protocol for communication between appliances and the server.

Domb proposed the smart home based on the Internet of Things. It incorporates all the basic features of home automation via communication and control of devices in [26].

This paper [27] by Gupta et al. presents the design and development of an effective system in managing home power consumption. The system primarily gathers information about the various home appliances and monitors and controls their performance for the most efficient power consumption scenario. The authors made use of a Power Line Communication (PLC) with a Power-Controlled Outlet Module (PCOM) to assist in the operation of the system. It consists of a home server connected to the module's network in a star topology arrangement.

However, the lack of an Internet connection is a significant drawback of the system as computation and control of home appliances cannot be made remotely. The lack of remote access to the system poses several problems, the most significant being that the system is less efficient in cases where emergency access is needed.

Ramljak [28] published the security analysis of open home automation bus systems and proffered excellent insight into the vulnerabilities and how to mitigate them effectively. The main focus was security analysis of the OpenHAB framework. It was done by traversing the security architecture and supported features that come with OpenHAB due to the apparent challenge in static code analysis of several used packages in OpenHAB. The author minimized vulnerabilities such as posting data to REST API, which can lead to leakage of user data by setting the X-XSS-Protection HTTP response to "1." The second approach the author used was to run OpenHAB behind the reverse proxy, which redirects client requests to the appropriate server, allowing access to OpenHAB runtime ports 80 (HTTP) and 443 (HTTPS).

Heimgaertner et al. [29] sought to help reduce energy costs and increase the comfort of living by adjusting room temperature according to schedules, rules, and sensor inputs. It utilizes the distributed OpenHAB Distributed Multiuser (DM) setup with extensions introducing user authentication, access control, and management tools for decentralized OpenHAB node deployments.

The limitation of this work is that the new OpenHAB DM is not compatible with existing mobile OpenHAB apps. However, this is mainly because the mobile apps do not currently supply the implementation of user credential functionality as well as provide authentication tokens to the REST API.

Celtek et al. [18] designed and implemented a low-cost, effective wireless home automation system with a general purpose application. The web-based application integrates with existing homes to control temperature, humidity, motion, and luminosity sensors along with the signal conditioning circuitry from user-inputted data. The home automation system consists of three key segments: the sensor nodes (SNs), the actuator nodes (ANs), and the center node (CN).

However, the absence of a mobile application interface reduces the ease of access to the system. The system does not define access roles and privileges for users of the systems, and this makes the system susceptible to unauthorized access. No security measures were taken to encrypt the information sent out by the system over the Internet.

Bhatt and Patoliya in [19] designed and implemented a scalable, cost-effective sensor wireless network for transforming the traditional home into a smart home. They deployed a heterogeneous sensor and actuator nodes based on wireless networking technologies for a home. The system has a scalable architecture, and thus, any number of home appliances can be controlled and monitored on the system. Additionally, the system can easily integrate existing wireless home appliances to present a single interface for the control of all home appliances. The authors employed the use of the ESP8266 WiFi module, and that posed a challenge being that it only provides a single point for analog input. That made it difficult to interface multiple analog sensors to the system. The authors, however, noted that integrating a separate Analog IO Expander circuit with the WiFi module would resolve the issue.

Song et al. [21] designed and implemented a wireless controllable power outlet system for home automation networks. Their power outlet module integrates a Zigbee radio that serves as an actuator node in the home automation networks. The smart home designed by the authors makes it easy to control various home appliances. However, a notable limitation of their research paper is the absence of a security standard for the wireless communication modules.

Karaca et al., in their paper [22], designed a smart home system employing the use of sensors and controller nodes connected to a custom local wireless network—the developed system functions in two primary ways: monitoring and control of home appliances. With the use of an embedded server, users can modify the database over an Internet connection. This embedded server is composed of an ARM microprocessor, security alarm horn, ESP8266 WiFi module, and an SD

card memory. The server is a bridge between the user interface and the sensor nodes/controllers. Server modules feed the user with required information by fetching data from the database, which is updated by the sensor nodes. This system does not implement any security standard to secure the wirelessly transmitted data to protect it from external hacks. Additionally, the authors recommend that future studies should be directed towards learning algorithms that enable adaptive control and machine learning control of parameters such as temperature increments in unit volume and time.

Due to fast steep growth in the usage and reliance on striking features of smart devices, Vikram et al. [23] proposed a methodology to provide a low-cost home automation system using WiFi. It embodies the concept of the internet-working of smart devices. The experimental rig involved the use of the ESP8266 WiFi module, ATmega microcontrollers (μ Cs), Single-Pole Double-Throw (SPDT) relays, Transceiver NRF2401+ RF, etc. One of the μ Cs was used as a hub, and all sensor readings are sent through the hub from the nodes via the RF modules.

A mobile application was developed with a user interface for controlling the home appliances. This paper failed to address the security aspect of the home automation system. The system does not provide access control levels for the user; hence, anyone can be granted permission to intrude the house. Also, since the paper made an emphasis on their system being WiFi-based, it failed to outline the security measures to handle external intrusions from hackers.

Gyory and Chuah in [24] proposed the design of an IoTOne solution that supports heterogeneous IoT devices and avoids both the security vulnerabilities and the limited device compatibility issues. The IoTOne solution supports heterogeneous IoT devices and provides security checking of IoTApp such that only IoTApp with robust codes would be hosted on the IoTOne App store. However, their system lacks support for COAP (Constrained Application Protocol) and DTLS (Datagram Transport Layer Security) features, which are essential for energy-efficient secure communications.

Shinde and Dube [4] developed an IoT-based smart energy management system to intelligently control the appliances wirelessly rather than just switching devices on or off. In this paper, the power consumption of appliances was measured in intervals of 30 minutes and sent to the Raspberry Pi. A significant drawback of the system is that it was designed for controlling only two appliances, i.e., electric fan and light bulb. Additionally, no security measure was devised, which is not safe due to the rigorous intrusion that is always attempted by hackers.

The paper published by Sowah et al. [30] provides a useful solution to the problem of the programmability and customizability of a modular home automation system. Three main features characterize the system: status, control, and automation of electrical and electronic devices. These three features are achieved through the robust hardware consisting of a router, system controller, wireless communication module, microcontroller module, fan controller module, lamp controller module, and IP camera. Primarily, the system uses Android application as a user interface to enable

communication between all system modules. There is a REST API server to facilitate communication between the database and the WebSocket. While the system works well in achieving the modular design for a home automation setting, it does nothing to secure the wireless network against data breaches and attacks [30, 31].

In these papers [3, 32], the authors designed an automated system for controlling lights and fans. A simple Graphical User Interface (GUI) was designed to make it easy for the end user to access data from a web interface. The system architecture comprised a Raspberry Pi, which serves as the Main Controlling Unit (MCU) and Input/Output Interface (IOI). Raspberry Pi was chosen as the MCU for good reasons—it is user-friendly and relatively cheap. With the Raspberry Pi as the central hub, the input can be fetched from the connected sensors and all the data are sent to the Raspberry Pi. However, the absence of a mobile application interface to the system reduces the ease of accessibility and enhanced functionality of the system. The system lacked security features to prevent network sniffing and session hijacking, which is very common with web-based applications.

All the reviewed papers were notable for successfully internetworking various home appliances and devices for monitoring and control purposes. Most of the papers reviewed were either web-based or Android applications that provided users with an interface to monitor and control the states of their home appliances. This was usually done through reliable and functional wireless communication modules. A few of the papers reviewed considered a security evaluation of the OpenHAB server software. Overall, most of the systems we considered proved to be low cost and user-friendly and proved to be ideal for monitoring and remote control purposes. A few of the systems maintained support for heterogeneous devices and implemented security vulnerability checks.

A closer look at the systems, however, revealed some gaps. Some of the systems failed to provide access control levels for users, making it easy for intruders to be granted access permissions. Other systems were found to be lacking support for multiple devices from different vendors due to compatibility issues. Some of the systems reviewed lacked remote control functionalities, and a few others did not implement any security mechanism for the wireless communication protocols. These gaps identified from the literature review informed our problem statement and possible solution approach in developing a secure wireless home automation system with an open home automation bus (OpenHAB 2) framework by leveraging and improving upon the default server architecture security features to enable privacy by design and control of home appliances via mobile and web apps.

3. System Design and Implementation

In the system design and implementation phase of the proposed solution to the identified gaps in the reviewed literature, both software and hardware tools were utilized. These tools are well suited for achieving our desired goal of building

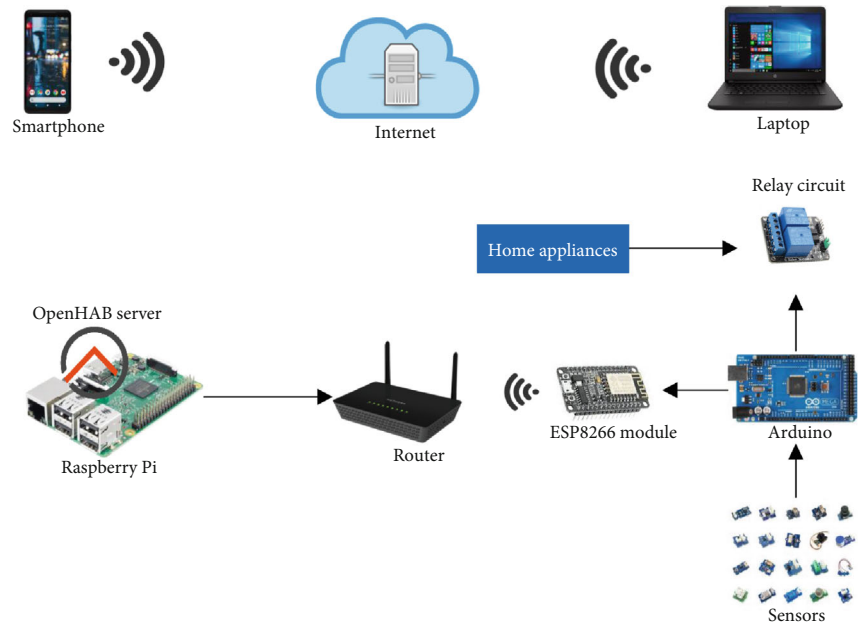


FIGURE 1: System architecture.

a secure wireless home automation system. In building the designed prototype, readily available circuit components were employed. The cost, efficiency, implication, and suitability of each of the components used in the prototype design were carefully weighed. The development tools used in this research paper were effective in helping to deliver the set objectives using the OpenHAB framework.

From the schematic shown in Figure 1, the system has two distinct parts: the wireless control system and the home network system. The smartphone and personal computer, acting as wireless controllers, are connected wirelessly to the network directly or over the Internet to enable remote control of home appliances as well as communication on updated states of the appliances. The Raspberry Pi microcontroller has the instance of OpenHAB software installed on it and acts as the central server for the system. It also hosts the MQTT server, which is a publisher- and subscriber-based protocol that allows multiple devices to communicate with each other over a wireless network [7]. The in-house router logs the IP addresses of the different devices connected to it and further acts as a central connection point for all the devices and appliances. The Arduino microcontroller communicates with the router via the ESP8266 module. It receives and broadcasts the state of the appliances via the MQTT protocol, where all clients can either subscribe or publish their status. An automated or user-queried action is then taken based on the subscribed device state. It is further connected to sensors that pick up signals from the surrounding environment. The Arduino microcontroller receives the command via the ESP8266 and effects the change via actuators (relay circuits and motors), which is essentially the connection point for the various home appliances.

The central server can be accessed by a user authorized by the system with *an email address and a password*. The connection can either be made through a web interface or

custom-designed Android application. It provides the user with the necessary data stored in the database. Upon gaining access to the central server, based on the information given, the user can then make queries (send commands). Figure 2 is the activity diagram of the overall system architecture that captures the flow of data and relates the communication and control of home appliances to achieve a smart home using OpenHAB 2. Similarly, Figure 3 shows how the system responds to the users' queries and instantaneously sends feedback. The responsiveness of the designed architecture was warranted because, in smart homes, appliances must respond to either switching commands or voice commands.

3.1. Circuit Design and Simulation. The circuit design and simulation using Proteus software was carried out first to ascertain the behavior of the various hardware modules when they are integrated. It included the power supply unit with the relays connected. The current and voltage sensor units are also displayed in Figure 4. The switching operation for energy management utilizes a single channel relay. The appliance switching control schematic using Proteus is shown in Figure 5.

The voltage sensor is used to monitor, calculate, and determine the voltage supply. This sensor can determine the AC or DC voltage levels of the connected appliance. The input to this sensor can be the voltage, whereas the output is the switches, an analog voltage signal, a current signal, an audible signal, etc. Some sensors provide sine waveform- or pulse waveform-like output, and others can generate outputs like AM (Amplitude Modulation), PWM (Pulse Width Modulation), or FM (Frequency Modulation). The measurement of these sensors can depend on the voltage divider.

On the other hand, a current sensor is a device that detects electric current flow and generates a signal proportional to that current flow. The generated signal may be

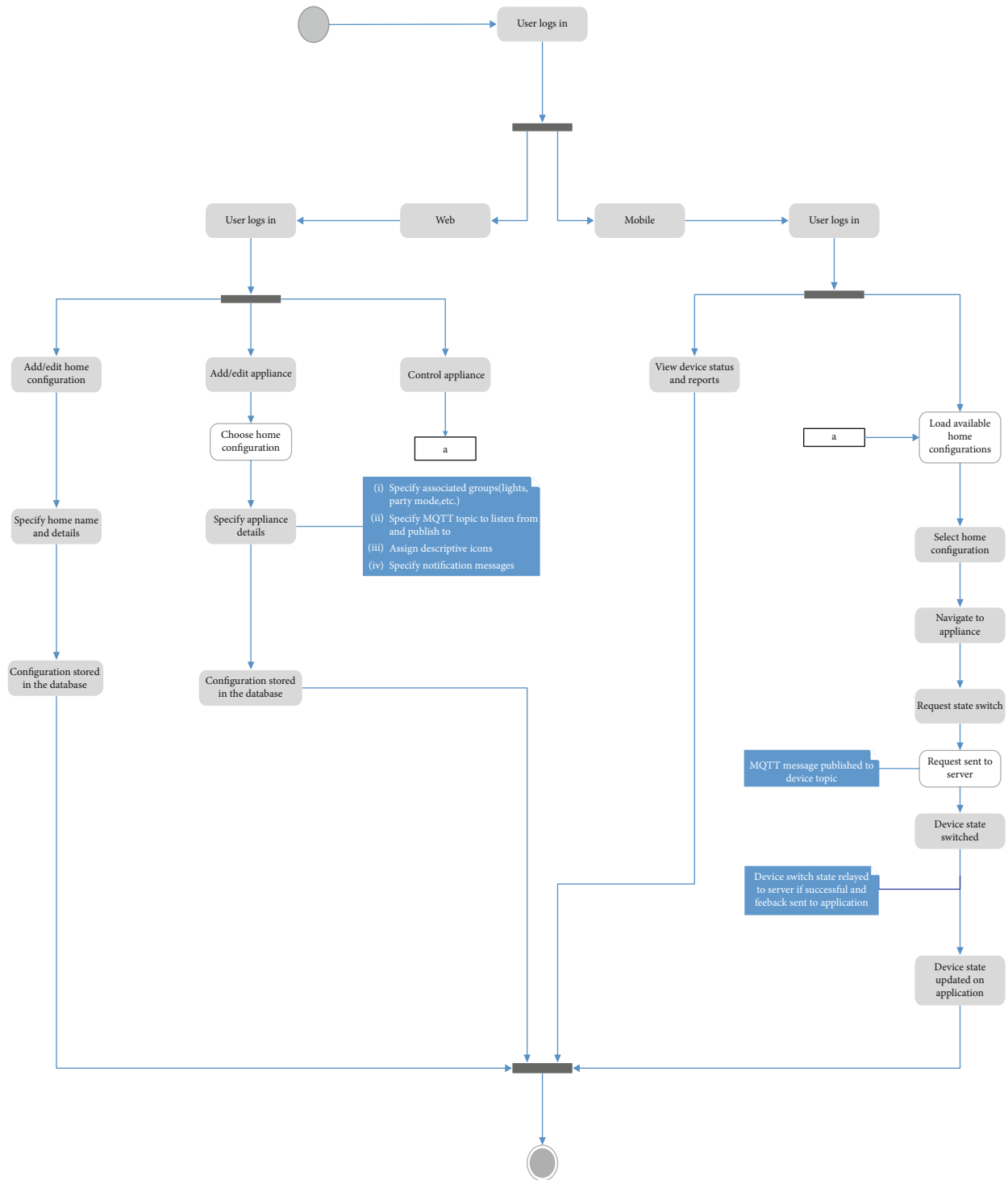


FIGURE 2: Activity flow diagram for the system architecture.

analog voltage or current, which can be utilized to display the measured current in an ammeter or stored for further processing and analysis in a data acquisition system or used for control of home appliances.

3.1.1. Main Hardware Components. Raspberry Pi. The model used in this research paper (Raspberry Pi 3 Model B) hosts a lightweight Linux distribution with OpenHAB installed. It

also has the Mosquitto MQTT broker installed. It communicates with the network by connecting either via Ethernet cable or wirelessly to the router.

Arduino Mega 2560. It acts as a control/logic unit for appliances that do not have a wireless switch. It saves the channel each appliance subscribes to or sensors publish to. It also interprets the MQTT messages sent and forwards the appropriate instructions to specified appliances.

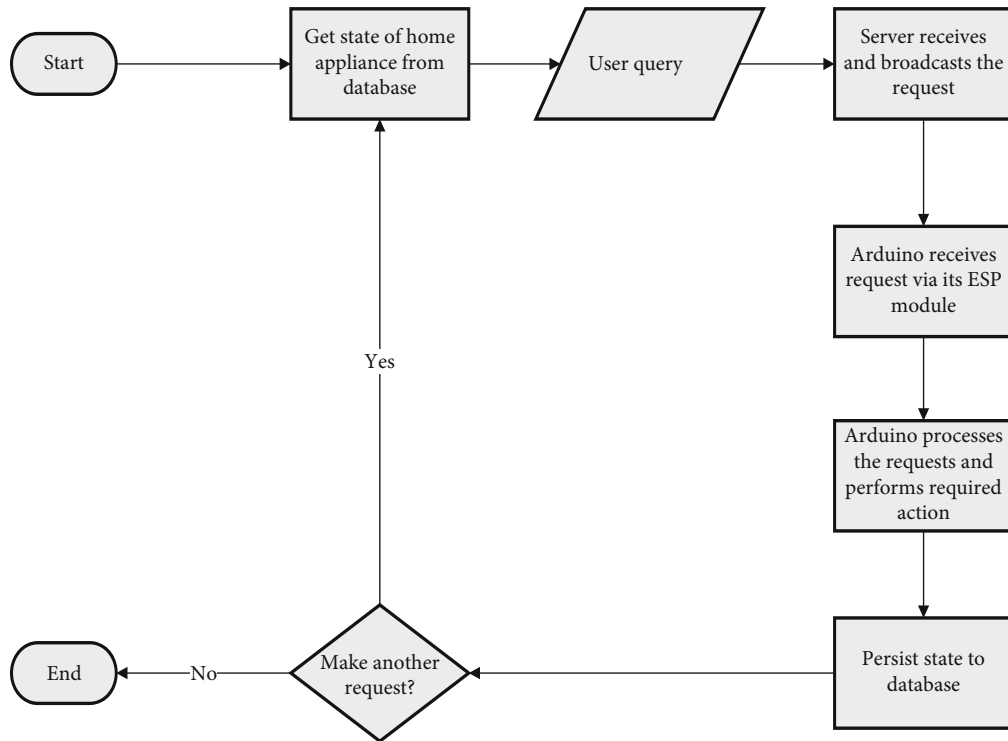


FIGURE 3: System flow diagram.

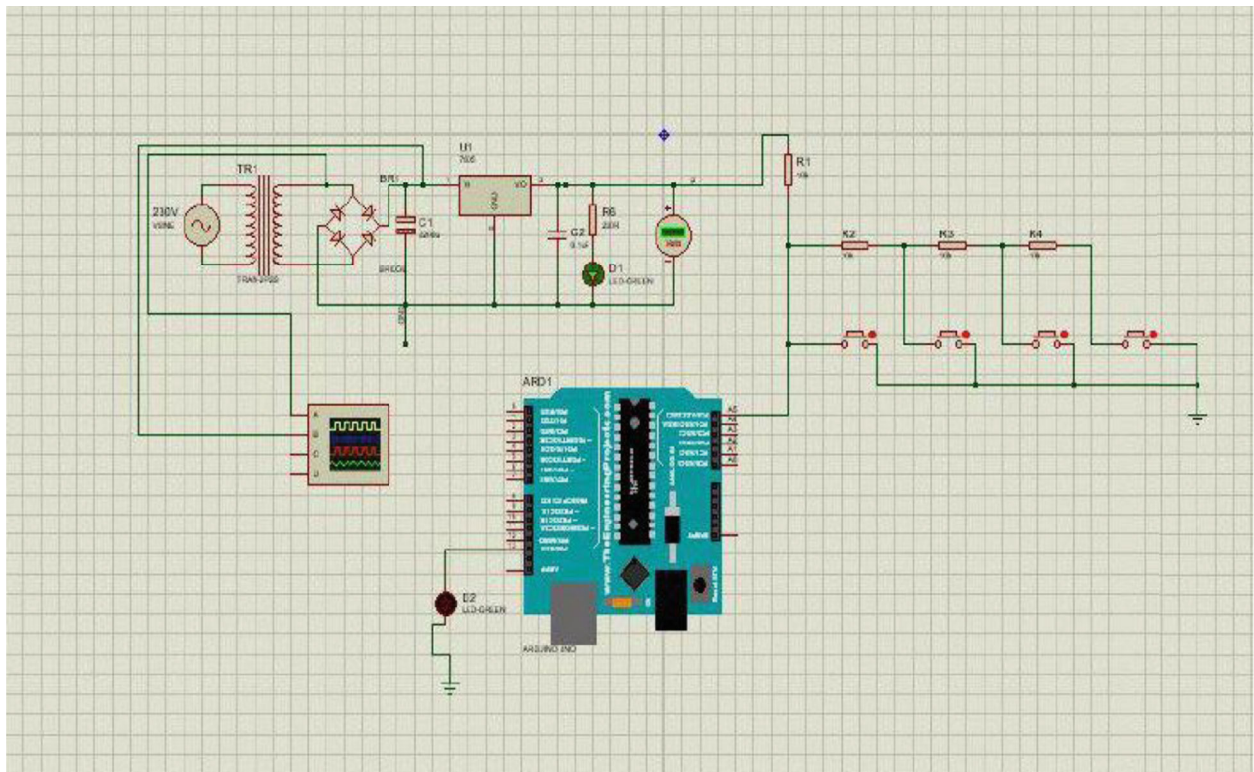


FIGURE 4: Circuit simulation for switching.

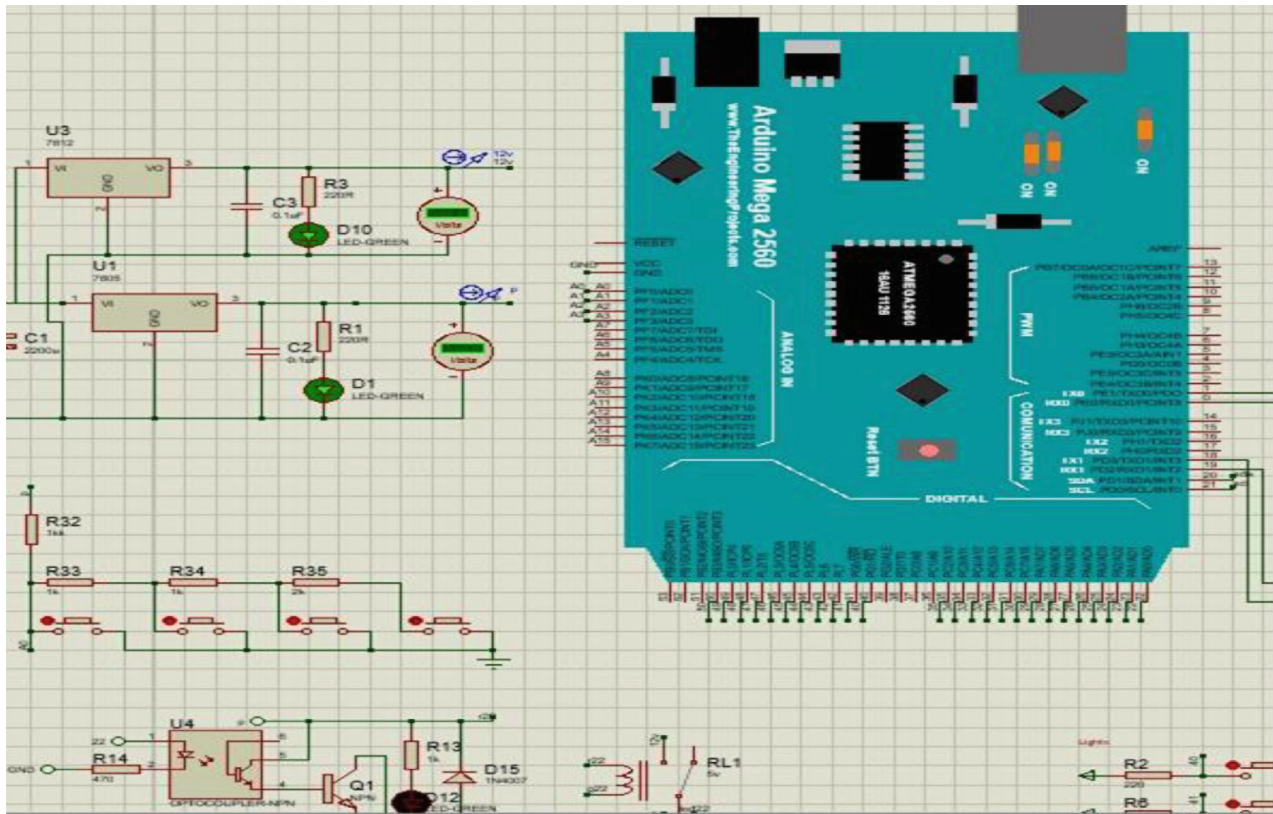


FIGURE 5: Appliance switching control circuit schematic.

NodeMCU (ESP8266) module. It enables wireless communication between the Arduino Mega 2560 and our Raspberry Pi. It also allows individual wireless switches within our system to communicate with the central server.

Relay modules. The relay module used was a 16-channel relay module. The 16-Channel 12V Relay Interface board allows for appliances to be controlled directly from the Raspberry Pi or Arduino microcontroller. This relay module allows a small-level signal to control any normal outlet or high-voltage product rated under 2000 watts.

Other hardware components are jumper wires, LEDs, capacitors, resistors, push buttons, step-down transformers, sensors (temperature, humidity), buzzer module (Piezo speaker), voltage regulator LD1117V33, and NPN bipolar transistors.

4. Hardware Programming and Application Programming

Every part of the system has some level of programming to enable flexibility and adaptable control to achieve the complete home automation system. The design software environments used in this paper for the hardware and application programming include the Arduino Integrated Development Environment (IDE), Android Studio, Visual Studio Code, Proteus Simulation Software, Fritzing, Etcher, PuTTY, and Vim. All code on any Arduino component was done with the Arduino IDE; the Android application was built with

Android Studio, the web view with Visual Studio Code, and configuration on our OpenHAB instance on the Raspberry Pi done over SSH using PuTTY with the Vim editor.

The different sections of the entire developed secure home automation system are as follows.

4.1. The Central Server. To set up the central server, a custom image of Linux distribution with an instance of OpenHAB called *openHABian* is loaded onto a memory card inserted into the Raspberry Pi using Etcher. The memory card is loaded, and the Raspberry Pi is then connected to the Internet to enable updates and initial setup. After 30 to 45 minutes, relative to Internet speed, the process is completed.

The openHABian instance is connected via SSH using PuTTY, as shown in Figure 6.

The configuration for the appliances is implemented in the corresponding files for each of the units below:

- (1) *Items.* This refers to the individual control devices and elements in our system. Each item can be given a name and assigned to groups with specific bindings.
- (2) *Sitemaps.* The sitemap is the interface the user interacts with when the OpenHAB mobile or web application is opened. The various button and information layouts can be controlled from the sitemap.
- (3) *Rules.* This is where the home automation logic is defined. Schedules and conditions required for actions to happen are all defined in this module.

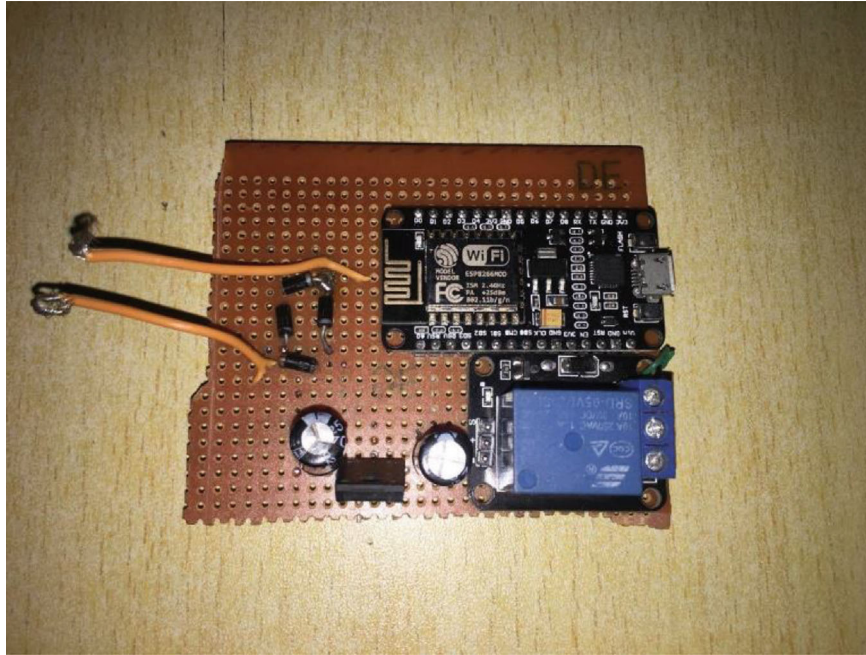


FIGURE 8: Wireless switch prototype implementation.



FIGURE 9: Raspberry Pi server and router setup.

mobile operating system market. Android remains the most popular mobile operating system in the world, having almost 75% of the global market share, according to StatCounter Global statistics [16]. The choice was made to build an Android application to cater for a considerable percentage of the users. While not wholly leaving out iOS in the development process, most of the customization work done on the OpenHAB application was mainly implemented on the Android app and written purely in Java.

In Figure 10, the various use cases for the application are shown. It summarizes the functions available to the home-

owner via the app. To demonstrate the functionality of the Android application, Figures 11 and 12 show authentication, which is a security feature on the smartphone application. For remote connections, even when not at home (local server), the remote connection to the OpenHAB cloud service is required. Otherwise, the user enters their details to be authenticated for access to the local server.

In Figures 13 and 14, the interface for the mobile application is shown. Figure 13 shows the main menu view for the sitemap of a given home. Figure 14 goes on to show the various rooms or spaces on the first floor of the home.

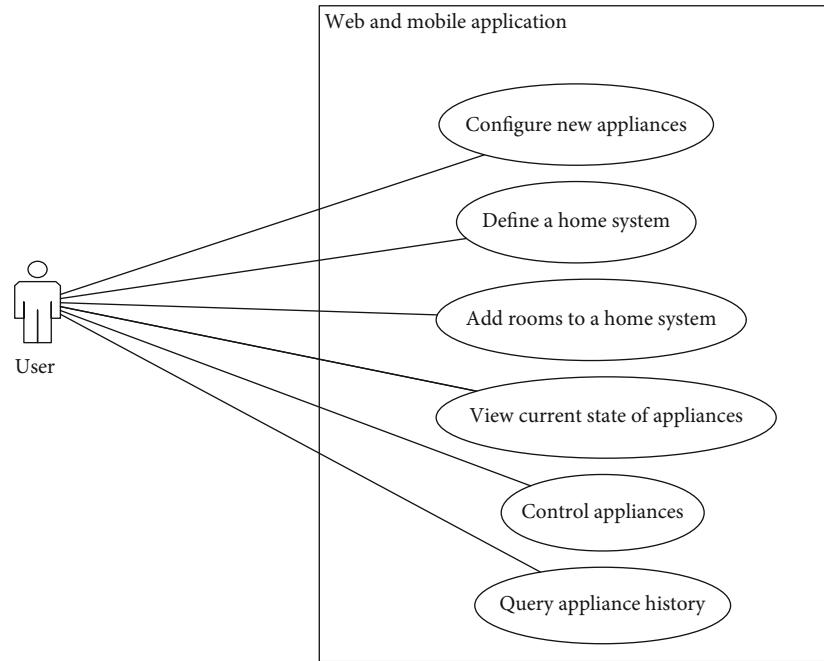


FIGURE 10: Application use case diagram.

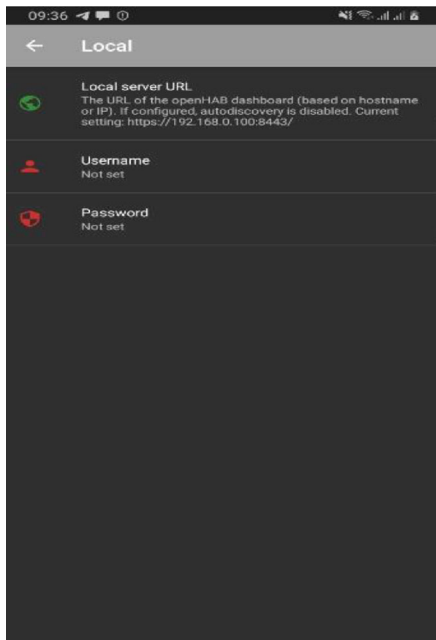


FIGURE 11: Local server connection.

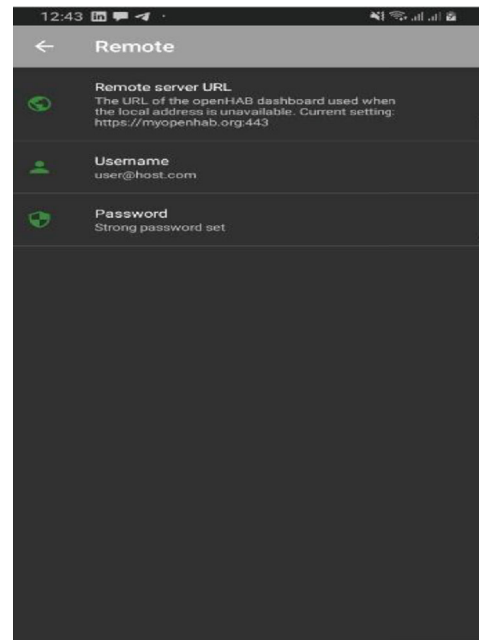


FIGURE 12: Remote server connection.

After selecting the “Corridor,” as shown in Figure 15, all connected appliances in that space, as well as their respective states, are retrieved and displayed to the user. Figure 16 shows the updated state after the light has been switched on.

These demonstrate the requirements of the smartphone application, which accents the importance of the home automation system and conclusively demonstrates monitoring, control, and some level of security.

4.5. Security Implementations by User Authentication and Authorization on OpenHAB. Security is an essential protection against anything that could pose a threat to a system. One of the biggest challenges with OpenHAB perhaps is the nonenforcement of an access control mechanism for its users, thereby making the security of the system dependent on the strength of the wireless network. We developed a simple authentication and authorization model to control access permissions to things on the network.

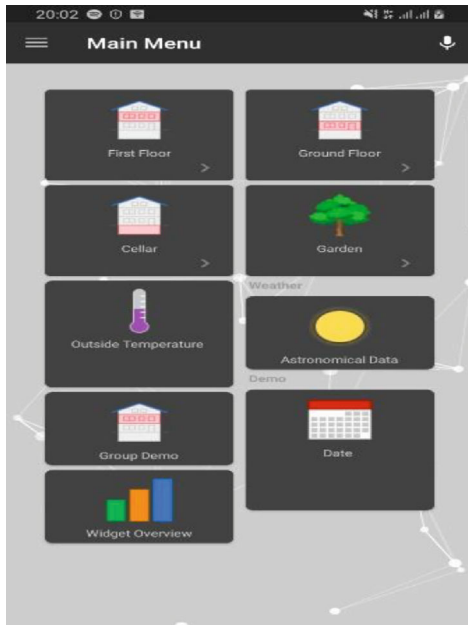


FIGURE 13: Main menu view.



FIGURE 15: Corridor light off.

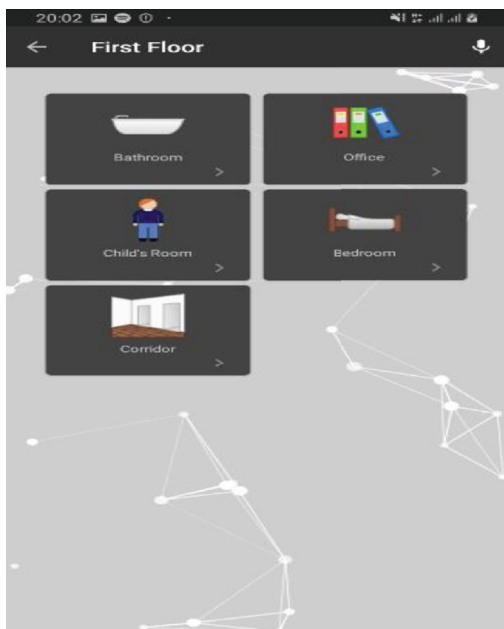


FIGURE 14: First floor view.

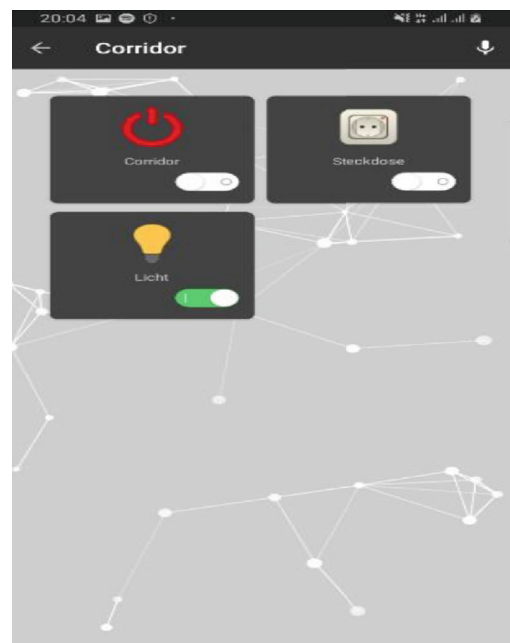


FIGURE 16: Corridor light on.

This was done using the JSON Web Token- (JWT-) based authenticator as a base model for all access control protocols. This base model depicted in Figure 17 is an abstraction of well-defined security policies and rules in our OpenHAB system.

JSON Web Tokens (JWT) are a tool that makes use of cryptography to ensure secure data communication between parties. There are two critical steps in using JWT securely in a web application: (1) sending them over an encrypted channel and (2) verifying the digital signature immediately upon receiving it. The asymmetric nature of the public key cryptography makes JWT signature verification possible. A public

key verifies a JWT was signed by its matching private key to enable data communications securely. No other combination of keys can do this verification, thus preventing impersonation attempts. JWT is a very modern, simple, and secure approach which extends for JSON Web Tokens. JSON Web Tokens are a stateless solution for authentication. Hence, no need to store any session state on the server, making it ideal for restful APIs, thereby making the server implementation lightweight and computationally efficient in the use of resources. Restful APIs should always be stateless, and

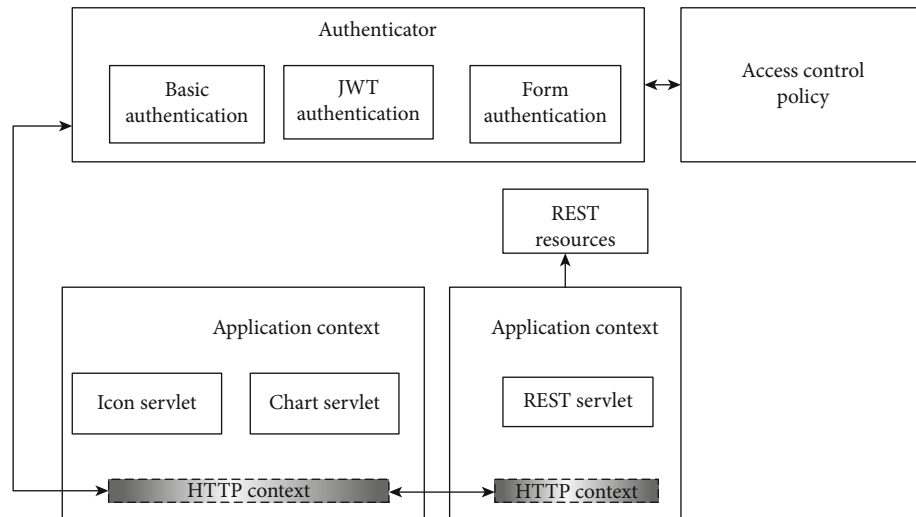


FIGURE 17: Block diagram of authenticators.

the most widely used alternative to authentication with JWT is just to store the user's login state on the server using sessions.

The first step in addressing the security challenges of OpenHAB was to analyze how data is moved through the network. Data flow through the network was monitored using Wireshark, a free and open-source packet analyzer. Based on the packet monitoring and tracing using Wireshark, it was discovered that data is transmitted in two ways, namely, via the cloud and the local OpenHAB host. Recognition of the fact that some devices usually do not need to connect to the Internet to set or reset their states was ascertained based on analysis, and that informed our design choices and protocols to use to establish secure communications. However, connection to the Internet is required in cases where the user might want remote access to things at home. The two ways of data transmission in the network required that we addressed them uniquely. The first scenario involves the case where OpenHAB establishes a connection through the Internet and opens up a lot of known vulnerabilities, one of which is that packets transmitted can be hacked and tampered with. In preventing this, we used the JWT authentication procedure, which we performed in five steps:

- (1) The client sends his/her details through the basic authenticator initialized on the web browser
- (2) The broker extracts the client details, and if it matches a registered client, a *JSON Web Token (JWT)* is generated. It is further appended to his unique ID along with the expiration date and digital signatures. The broker relays the JWT to the client
- (3) The client attaches the JWT on any subsequent request to the server
- (4) For every request, the server extracts the JWT and verifies the digital signature appended to it. If it is deemed valid, the username and password are used to perform authorization on the system for use

- (5) If the JWT is expired, the broker requests user details through basic authentication, and if they are valid, it generates and serves another valid JWT

The second case where OpenHAB communication is done locally between home appliances and the OpenHAB host appeared to be of lesser risk. This is because communication is usually through a wireless network encrypted with AES. Breaking the AES is almost computationally impossible, especially for the AES-256 encryption enforced on the developed system. The sequence diagram of the security implementation is depicted in Figure 18.

4.6. Hardware and Software Integration. The device controller, in two modes, controls the devices. OpenHAB serves to handle the remote mode while the switching circuit developed handles the manual mode. The DHT11 Temperature and Humidity sensor features three pins: VCC, DATA, and GND. The VCC is connected to the 5V, GND to GND of Arduino, and DATA to the analog pin of the Arduino microcontroller. On the Arduino IDE, we imported the *DHT.h library*, which enabled us to convert the analog value read into real-world temperature and humidity values. This data is published via the ESP8266 module to the server on the Raspberry Pi and stored securely for remote monitoring.

5. Testing and Results

The developed system integrated all items to a single switching board to show that the states of devices could be controlled remotely with improved security. From tests conducted, the developed system failure rate is exceptionally minimal, thus offering a high level of security in home automation, and the set objectives of secure wireless switching and internetworking of home appliances and subsequent control were achieved.

5.1. Android Application Testing. The first approach to testing the Android application was to check for stability across

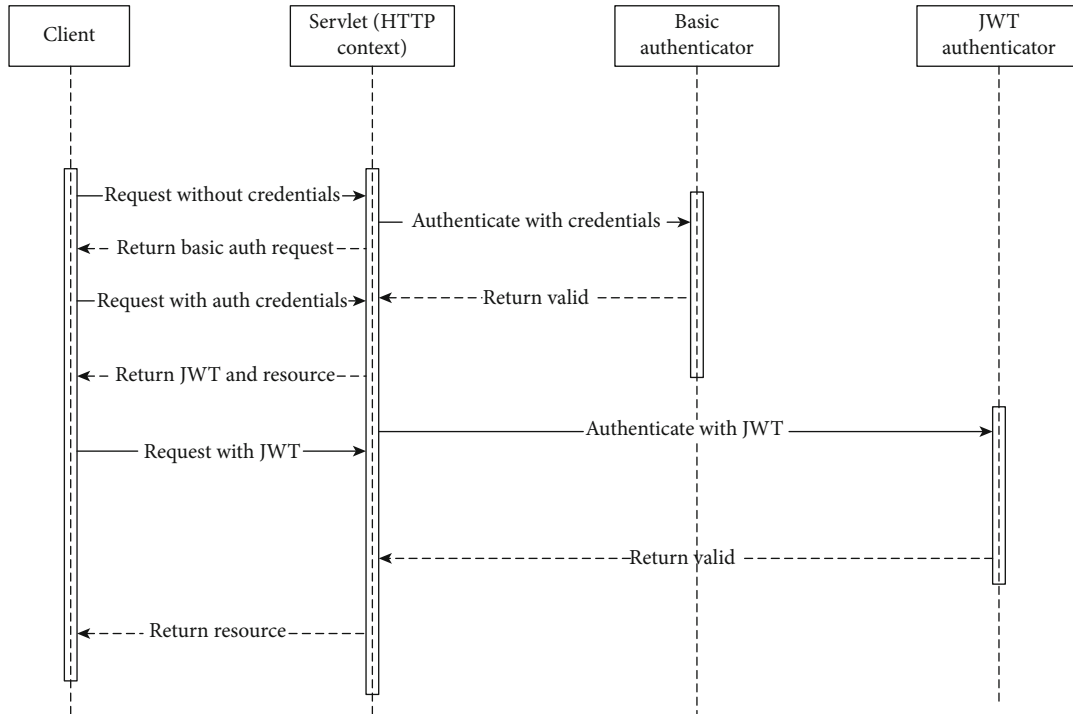


FIGURE 18: Sequence diagram of the security implementation.

a range of Android devices. After looking into the various Android operating systems to determine the relevant ones to test our mobile application, the test focused on a limited set of most popular devices and Operating System (OS) versions since it would be impractical to test out the application on every Android device on the market.

The application was run on one high-end Android device, the Samsung Galaxy S9+, and a midrange device, the Samsung Galaxy A5. The application performance on both devices was similar. No visible performance lags or GUI responsiveness issues.

During mobile application development, the JUnit, a testing framework, was used to test all modules within the application. The test helped us to fix some significant compatibility issues across the different Android operating systems.

Tests performed considered the effect the application running has on battery life. The application had minimal impact on the battery, using less than 1% of battery on standby and about 10% on continuous 5-hour usage. Comparing this with the average of all data-enabled applications, the net battery savings recorded was about 15%. Data consumption was also minimal, with the application consuming only about 5 MB within a 5-hour period of close to constant usage after deployment.

5.2. Web Service Testing. In testing the web service, PostMan and HttpMaster were used. Both applications support dynamic parameters, data validation, and response data. The various requests yielded the expected response codes, hardly ever reporting failures even on very slow and unstable networks.

5.3. OpenHAB Testing. The testing of the OpenHAB implementation focused on the overall system latency. Various switching commands were issued out to the system. On a local network, it was observed that switching requests and responses were much faster than when on the cloud service. Since a guarantee cannot be made for great user Internet speed, this issue was addressed by prioritizing the local server over the cloud server to reduce network latency.

5.4. OpenHAB Security Tests. OpenHAB makes use of bindings, which is a logical software piece that links a thing to OpenHAB. The data transmitted through the REST API calls on the bindings can be eavesdropped upon if the packets through the network are inspected. Wireshark was used to inspect packets and understand how the different bindings transmit data.

Internal communications between things in the network and OpenHAB framework are through a wireless network, and it is often encrypted with AES. An eavesdropper can get the transmitted data if they can decrypt the AES signal data transmitted. That is, in itself, computationally not feasible. Security, in this case, as we tested, was dependent on the strength of the wireless network.

Another communication approach was where OpenHAB was communicating with a remote server in the cloud. In addressing this security challenge, a JWT-based authenticator was implemented in Java. Testing directly on the OpenHAB framework was not an easy task, so we first demonstrated the implementation as an independent deployment. In this independent deployment, we ran the system against various intrusion tests and packet/session hijacking tests using industry-standard tools for which

TABLE 2: Hardware system tests and results.

Test	Expected result	Result	Remarks
Start with the hub, and OpenHAB service configuration is running on the hub without an Internet connection	Initialize the database connection, start OpenHAB service, and connect to cloud service if available. Establish a connection with the MQTT server and update device states	Successful	Start-up takes about 45 seconds on average (out of 20) to start up Raspberry Pi and OpenHAB service
Connect a LAN or WAN to the hub while OpenHAB service configuration is running	Trigger cloud connects service to establish a connection with the cloud server and sync device states	Successful	Cloud service worker detects Internet connectivity, and cloud loads configurations with a connection 10 seconds
Stop the OpenHAB hardware application or power off Raspberry Pi while OpenHAB service is still running	Raise <i>ForcedTermination</i> event in OpenHAB Control Panel logs and publish to <i>ForcedExceptions</i> MQTT topic	Failed some of the time	Failed some of the time. The cause is a delay in publishing before power is cut
Create a new "Home" location in the OpenHAB Control Panel	The location should be added to the local database and posted to and made available on the OpenHAB server	Successful	A new location is immediately available
Add a new appliance in the OpenHAB Control Panel	The appliance with its properties should be added to the local database with OpenHAB service refreshed to show the update as well as send push notification to registered users	Successful	A new appliance was immediately available. The push notifications, however, worked 99% of the time
System vulnerability test	No breaches or vulnerabilities during test runs locally at the client as well as remotely		Firewall bypass testing was a success. The system withstood DNS-level attacks, e.g., switching and routing tests

TABLE 3: Software system tests and results.

Test	Expected result	Result	Remarks
Start the OpenHAB mobile application	Verify OAuth token with the OpenHAB local server and cloud service, while the splash screen is shown, display login state if the token is invalid. Initialize the server connection, synchronize the current device states, and display the main page	Successful	With Internet connectivity, 1.5 secs is required from splash screen to the main screen with a local server and 3.5 secs for an active cloud configuration device information and states
Switching between the local server and the cloud service	Switch from the local server to the cloud service when out of range of the local server and switching back	Successful	Also, prioritizes local server when back in range without compromising connection or device states
Create a new OpenHAB account	Save the user account details and issue the OAuth token	Successful	
Request from OpenHAB mobile to cloud service on OpenHAB web	Verify the OAuth token if it passes, fetches the required data from the database, updates, and returns to OpenHAB service	Successful	One out of about 100 tries does not update on the app after the request is successful. Refreshing the app page, however, shows the updated state
Interact with a hardware control for a device on the OpenHAB Control Panel	Effect change on the hardware device and send an update via MQTT to the OpenHAB server for relaying to all other clients then update cloud service	Successful	
Interact with a control for the power switch of an appliance in OpenHAB mobile	Send change via MQTT to OpenHAB cloud service for relaying to the OpenHAB Control Panel (Raspberry Pi)	Successful	Unless there is no active connection to either the local or cloud server, which is updated when there is a connection

TABLE 4: System security tests and results.

Test	Process	Observation	Remarks
Wireless network penetration testing	<p>Used industry-standard penetration testing tools and frameworks, including Aircrack-ng. The remote penetration testing device was placed within the network. The wireless network audit began with a full sweep of the 2.4 GHz wireless frequencies, where numerous busy networks were found</p>	<p>1. It uses the preshared key for network access by sniffing the network while forcing an existing client off the network 2. Capturing of a WPA2 handshake attack 3. Nonidentification of clients connecting to the hidden SSIDs during the audit period, and therefore, it was not possible to unmask them</p>	<p>A scan was made for Cisco Discovery Protocol (CDP) traffic, which would have disclosed further information about the network. CDP was not found to be running across the public guest network, and VLAN hopping was unsuccessful Risk level: extremely low</p>

the system withstood. The Aircrack-ng software, as a complete suite of tools to assess WiFi network security, was utilized. This tool focuses on different areas of WiFi security, such as monitoring of packet data capture and export of data to text files for further processing, attacking the network using replay attacks, deauthentication, fake access point creation, and packet injection with testing and checking of WiFi cards and driver capabilities. The Aircrack-ng software provides capabilities for cracking passwords or passphrases on Wired Equivalent Privacy (WEP) and Wireless Protected Access (WPA) protocol implementation.

The results of the various tests for the hardware system and integration, as well as software system and integration, are captured in Tables 2 and 3. Table 4 presents the results of the system security tests that were performed using industry-standard tools.

To test all cases in a simulated home environment, a demo house was constructed and fitted with the required hardware. Figure 19 shows the constructed demo house having a model garage, two bedrooms, a kitchen, a corridor, and a dining area with fitted hardware.

All the rooms have an LED serving as an energy-efficient light bulb, relays, and switches for turning lights on and off remotely via the mobile and web applications developed.

Figure 20 also shows the control circuit with the Arduino Mega fitted to enable control of connected devices.

The final setup with enabled connection to the central server is shown in Figure 21.

6. Conclusion

Most homeowners in developing countries lack complete and total control over their homes. They are not able to access vital home automation features such as control and monitoring of home appliances, low-cost security, and efficient energy usage by implication. The OpenHAB 2 protocol is relevant to automating the home while interconnecting appliances based on its flexibility and capability for full customization, to achieve the goals of (1) providing a secure control mechanism for home appliances via a mobile or web application; (2) improving the security of connected home appliances through the system's inbuilt intrusion detection, alarm, and wireless communication data encryption standards; (3) meeting the essential energy management requirement of the household by providing a means to monitor and remotely turn off unnecessary active appliances to conserve energy and reduce electricity bills; (4) offering a wirelessly controlled switch for all home appliances; and (5) making a significant contribution to the existing body of knowledge on secured home automation. This paper presented a secure wireless home automation system that has been designed and implemented with the OpenHAB 2 home automation software framework to meet the set goals. A power supply circuit was designed and implemented for the microcontrollers while another circuit was created for the wireless switching and control of appliances. The OpenHAB server was set up on the Raspberry Pi, and the Arduino was configured to communicate with the OpenHAB server for home automation tasks. Both mobile and web applications



FIGURE 19: Constructed demo house with LEDs integrated.

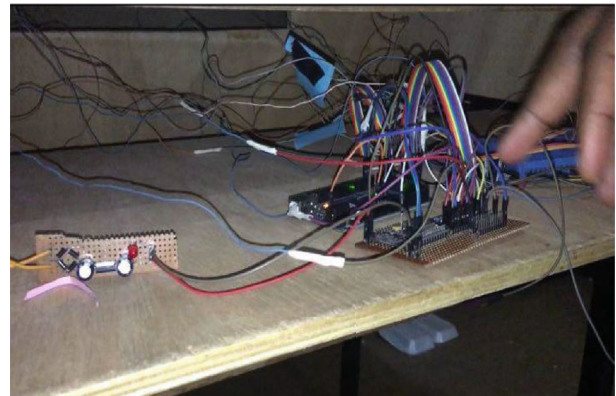


FIGURE 20: Compartment for housing demo house circuitry.



FIGURE 21: Demo house with all LEDs, motors, and sensors integrated.

were developed to control and view the status of home appliances. Also developed was a secure wireless network system enabling communication between the home appliances and the OpenHAB server. The prototype integrated all items to a single switching board as a proof of concept that the states of the device could be controlled and monitored remotely with improved security using JSON Web Tokens. Future research directions could be geared towards integrating the switching circuitry within the real-world appliances to trigger the switching internally instead of just controlling the power

supply to the appliance. Also, learning algorithms to enable adaptive control and machine learning could be implemented to handle sensor parameters automatically (e.g., from temperature, smoke, and other detector systems). Also, future project directions for enhancements could be geared towards the use of machine learning techniques to learn from the homeowner's behavior to optimize the switching of appliance states for energy savings using various optimization techniques.

Data Availability

The data used for software testing and penetration testing are available upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest in this research work and subsequent paper publication.

Acknowledgments

The authors would like to express their profound and earnest gratitude to the entire staff of the Computer Engineering Department, University of Ghana, for the invaluable knowledge and insightful comments on earlier drafts of this manuscript.


References

- [1] J. Jin, Y. Wang, K. Zhao, and J. Hu, "Development of remote-controlled home automation system with wireless sensor network," in *2008 Fifth IEEE International Symposium on Embedded Computing*, pp. 169–173, Beijing, China, 2008.
- [2] A. A. Zaidan, B. B. Zaidan, M. Y. Qahtan et al., "A survey on communication components for IoT-based technologies in smart homes," *Telecommunication Systems*, vol. 69, no. 1, pp. 1–25, 2018.
- [3] M. D'Souza, N. Wilfred, R. Pereira, T. Rayen, and A. Telgote, "Home automation using Internet of Things," in *2017 International Conference on Energy, Communication, Data Analytics, and Soft Computing, ICECDS*, pp. 559–561, Chennai, India, 2018.
- [4] M. Shinde and R. R. Dube, "IOT Based Energy Monitoring and Management System for Smart Homes," *International Journal of Recent Trends in Engineering and Research*, vol. 4, no. 1, pp. 287–295, 2018.
- [5] M. Daneshvar, M. Pesaran, and B. Mohammadi-Ivatloo, "Transactive energy in future smart homes," in *The Energy Internet*, pp. 153–179, Elsevier, 2018.
- [6] R. Teymourzadeh, S. A. Ahmed, K. W. Chan, and M. V. Hoong, "Smart GSM based home automation system," in *2013 IEEE Conference on Systems, Process & Control ICSPC*, pp. 306–309, Kuala Lumpur, Malaysia, 2013.
- [7] R. K. Kodali, V. Jain, S. Bose, and L. Boppana, "IoT based smart security and home automation system," in *2016 International Conference on Computing, Communication, and Automation ICCCA*, pp. 1286–1289, Noida, India, 2016.
- [8] W. Ejaz and A. Anpalagan, "Internet of Things for smart cities: overview and key challenges," in *Internet of Things for Smart Cities*, pp. 1–15, Springer, 2019.
- [9] S. Mahmud, S. Ahmed, and K. Shikder, "A smart home automation and metering system using internet of things (IoT)," in *2019 International Conference on Robotics, Electrical and Signal Processing Techniques*, pp. 451–454, Dhaka, Bangladesh, 2019.
- [10] W. Li, T. Logenthiran, V. T. Phan, and W. L. Woo, "A novel smart energy theft system (SETS) for IoT-based smart home," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5531–5539, 2019.
- [11] W. A. Jabbar, T. K. Kian, R. M. Ramli et al., "Design and Fabrication Of Smart Home with Internet of Things enabled automation system," *IEEE Access*, vol. 7, pp. 144059–144074, 2019.
- [12] TheAmbient, *Around the world in smart home trends*, The Ambient, 2018.
- [13] O. Veras, *Smart cities in Africa: Nairobi and Cape Town*, Africa Business Insight, 2017.
- [14] D. Hendricks, *The history of smart homes*, IoT Evolution World, 2014.
- [15] Research and Markets, *Global home automation and control market 2014-2020 - lighting control, security & access control, HVAC control analysis of the \$5.77 billion industry*, 2015, <https://www.reuters.com/article/research-and-markets-idUSnBw195490a%2B100%2BBSW20150119>.
- [16] *Mobile operating system market share worldwide: concatenated 2015 to 2019*, Statcounter Global Stats, 2019, <https://gs.statcounter.com/os-market-share/mobile/worldwide/2019>.
- [17] January 2017, <https://www.openhabfoundation.org/2017/openhab2>.
- [18] S. A. Celtek, M. Durgun, and H. Soy, "Internet of Things based smart home system design through wireless sensor/actuator networks," in *2017 2nd International Conference on Advanced Information and Communication Technologies (AICT)*, pp. 15–18, Lviv, Ukraine, 2017.
- [19] A. Bhatt and J. Patoliya, "Cost-effective digitization of home appliances for home automation with low-power WiFi devices," in *2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics AEEICB*, pp. 643–648, Chennai, India, 2016.
- [20] J. Kizza and F. M. Kizza, "Access control, authentication, and authorization," in *Securing the Information Infrastructure*, pp. 180–208, IGI Global, 2008.
- [21] G. Song, F. Ding, W. Zhang, and A. Song, "A wireless power outlet system for smart homes," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 4, pp. 1688–1691, 2008.
- [22] S. Karaca, A. Sisman, and I. Savruk, "A low-cost smart security and home automation system employing an embedded server and a wireless sensor network," in *2016 IEEE 6th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, pp. 73–77, Berlin, Germany, 2016.
- [23] N. Vikram, K. S. Harish, M. S. Nihaal, R. Umesh, A. Shetty, and A. Kumar, "A low-cost home automation system using WiFi based wireless sensor network incorporating Internet of Things (IoT)," in *2017 IEEE 7th International Advance Computing Conference (IACC)*, pp. 174–178, Hyderabad, India, 2017.
- [24] N. Gyory and M. Chuah, "IoTOne: integrated platform for heterogeneous IoT devices," in *2017 International Conference on Computing, Networking and Communications (ICNC)*, pp. 783–787, Santa Clara, CA, USA, 2017.
- [25] R. A. Atmoko, R. Riantini, and M. K. Hasin, "IoT real time data acquisition using MQTT protocol," *Journal of Physics: Conference Series*, vol. 853, article 012003, 2017.

- [26] M. Domb, "Smart home systems based on Internet of Things," in *IoT and Smart Home Automation [Working Title]*, IntechOpen, 2019.
- [27] A. Gupta, A. Mudgal, C. Jayaraj et al., "Smart home device and energy management systems," in *2011 Annual IEEE India Conference*, pp. 1–5, Hyderabad, India, 2011.
- [28] M. Ramljak, "Security analysis of Open Home Automation Bus system," in *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1245–1250, Opatija, Croatia, 2017.
- [29] F. Heimgaertner, S. Hettich, O. Kohlbacher, and M. Menth, "Scaling home automation to public buildings: a distributed multiuser setup for OpenHAB 2," in *2017 Global Internet of Things Summit (GIoTS)*, pp. 1–6, Geneva, Switzerland, 2017.
- [30] R. A. Sowah, A. R. Ofoli, M. K. Tetteh, R. A. Opoku, and S. K. Armoo, "Demand Side Management of Smart Homes Using OpenHAB Framework for Interoperability of Devices," in *2018 IEEE 7th International Conference on Adaptive Science & Technology (ICAST)*, pp. 1–8, Accra, Ghana, 2018.
- [31] R. A. Sowah, K. O. Apeadu, A. Ofoli, K. Koumadi, A. Acakpovi, and S. K. Armoo, "Interoperability of heterogeneous appliances in home automation using the AllJoyn framework," in *2018 IEEE 7th International Conference on Adaptive Science & Technology (ICAST)*, pp. 1–9, Accra, Ghana, 2018.
- [32] S. Dey, A. Roy, and S. Das, "Home automation using Internet of Thing," in *2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 1–6, New York, NY, USA, 2016.

Research Article

Application of Artificial Intelligence Techniques in Predicting the Lost Circulation Zones Using Drilling Sensors

Abdulmalek Ahmed,¹ Salaheldin Elkatatny ,¹ Abdulwahab Ali,² Mahmoud Abughaban,³ and Abdulazeez Abdurraheem¹

¹College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum & Minerals, 31261 Dhahran, Saudi Arabia

²Center of Integrative Petroleum Research, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

³Saudi Aramco, EXPEC ARC, Saudi Arabia

Correspondence should be addressed to Salaheldin Elkatatny; elkatatny@kfupm.edu.sa

Received 13 April 2020; Revised 1 September 2020; Accepted 9 September 2020; Published 22 September 2020

Academic Editor: Rafael Morales

Copyright © 2020 Abdulmalek Ahmed 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.

Drilling a high-pressure, high-temperature (HPHT) well involves many difficulties and challenges. One of the greatest difficulties is the loss of circulation. Almost 40% of the drilling cost is attributed to the drilling fluid, so the loss of the fluid considerably increases the total drilling cost. There are several approaches to avoid loss of return; one of these approaches is preventing the occurrence of the losses by identifying the lost circulation zones. Most of these approaches are difficult to apply due to some constraints in the field. The purpose of this work is to apply three artificial intelligence (AI) techniques, namely, functional networks (FN), artificial neural networks (ANN), and fuzzy logic (FL), to identify the lost circulation zones. Real-time surface drilling parameters of three wells were obtained using real-time drilling sensors. Well A was utilized for training and testing the three developed AI models, whereas Well B and Well C were utilized to validate them. High accuracy was achieved by the three AI models based on the root mean square error (RMSE), confusion matrix, and correlation coefficient (R). All the AI models identified the lost circulation zones in Well A with high accuracy where the R is more than 0.98 and RMSE is less than 0.09. ANN is the most accurate model with $R = 0.99$ and $RMSE = 0.05$. An ANN was able to predict the lost circulation zones in the unseen Well B and Well C with $R = 0.946$ and $RMSE = 0.165$ and $R = 0.952$ and $RMSE = 0.155$, respectively.

1. Introduction

The demand for drilling high-pressure, high-temperature (HPHT) wells has become more significant in the petroleum industry. HPHT wells are known for bottom-hole pressures of more than 10,000 psi and bottom-hole temperatures of more than 300°F [1]. The main advantages of drilling HPHT wells are increasing oil production and improving economic success [2]. Drilling these wells involves some challenges and difficulties, mainly the appropriate tools for formation evaluation. These include many areas such as HPHT cement integrity, modified testing procedures and equipment, battery technology, proper zonal isolation, elastomers, alternative sealing agents, electronics/sensors and selection of drilling mud, design of drill string, and bits [3].

Drilling HPHT wells can result in many problems that delay the drilling operation and impact the cost. Loss of circulation is a common problem in drilling these wells. The partial or entire loss of the drilling mud from the wellbore to the formation is called loss of circulation or loss of return [4]. The losses will occur when a path for the flow exists, while the pressure inside the well is greater than the formation pressure [5].

Loss of circulation increases the nonproductive time spent on mitigating the losses [6], besides increasing the total drilling cost due to loss of drilling mud, which represents, in some cases, 40% of the total cost. The oil and gas industry reported more than \$12 billion in the cost of drilling materials and fluids in 2018 [7]. Loss of circulation leads to poor hole cleaning due to the reduction of mud level in the

borehole, which decreases its ability to transfer the cutting outside the wellbore [8]. The decrease in the mud level might reduce the hydrostatic pressure and cause a kick or blowout if the wellbore pressure became less than the formation pressure [9].

A range of methods has been used to overcome the circulation loss. The first method is adjusting the properties of drilling mud to reduce the equivalent circulation density (ECD) and consequently decreasing the quantity of the lost drilling mud [10]. The second method is pumping the lost circulation material (LCM) to seal and plug the losses [11]. Nevertheless, these methods are time-consuming and very expensive [12].

To minimize loss of return, it is essential to identify the lost circulation zones. Although various approaches are available, such as ECD, temperature profile, and resistivity [13, 14], nevertheless, some of these approaches are impractical either due to the high cost or lack of technology or owing to inaccurate prediction of the thief zones.

2. Artificial Intelligence (AI)

Artificial intelligence (AI) allows computers to perform tasks that require human intelligence. According to Mohaghegh et al. [15], AI is aimed at building a model or an algorithm, which enables machines to perform duties that need knowledge, understanding, and experience when performed by humans. A broader definition includes problem-solving, language perception, and conscious and unconscious processes [16]. AI is also known as a subfield of computer science involving the use of computers in tasks, which usually needs reasoning, knowledge, learning, and understanding abilities.

2.1. Artificial Neural Network (ANN). ANN is an information-processing system, which attempts to imitate the performance features of the biological nervous system. The network is adapted as a computer model, which can advance transformations, associations, or mappings between data [17]. The feature of ANN is that it does not require any physical phenomenon that explains the system under study [18]. Any nonlinear complex function can be approximated by ANN to make a relationship between input and output parameters.

According to Ahmed et al. [19], an artificial neural network is made up of many components such as neurons, hidden layers, transfer function, learning function, training function, and epoch size. Neurons are components that have specific input/output, and they are connected to form a network of nodes that makes the neural networks [20]. Weights and biases are used to handle the input parameters to find a relationship between the neurons and the source, so the performance of the network depends on the selection of those weights and biases. Also, the performance of the estimation model relies on the choice of the number of hidden layers, training function, and number of neurons [21].

2.2. Fuzzy Logic (FL). Fuzzy logic (FL) is a method of reasoning where the rules of deduction are estimated rather than precise. FL is valuable for handling information that is

incomplete, inaccurate, or irresponsible. FL is closely similar to the theory of fuzzy groups that belongs to a set of objects with boundaries in which membership is a problem of degree [22].

The fuzzy system is typically used to characterize uncertainty, which is due to the imprecision of the data or insufficient input variables that have an essential effect on the results. A property or an item can be defined by categorizing it under one of the different noncrisp groups and also a degree of membership for every group [23]. The fuzzy set theory proposes that a truth value that is between 0 and 1 should be added when working with noncrisp variables.

A membership function is used to define the relationship between a truth value and its variable. It has a value between 0 and 1, and that describes the “degree” of membership [24]. The membership functions can be represented by different functions such as sigmoid, Gaussian, trapezoidal, or straight lines [25]. When set membership had been defined again in this method, you can explain a reasoning system based on techniques for relating distributions [26].

The fuzzy inference system (FIS) is the procedure of creating a formulated mapping from an input to an output. The system contains logical processes, a set of “if-then” rules, and formulating membership functions. FIS is composed of five main parts: fuzzification interface, rule base, database, decision-making unit, and defuzzification interface, as shown in Figure 1. At first, the fuzzification interface transfers the input data into degrees of a match with linguistic values. Then, the rule applies a number of fuzzy “if-then” rules. Databases are used in the rules for membership function, and the decision-making unit is utilized for the operations of inference. Finally, the defuzzification interface transfers fuzzy output to crisp results.

According to Jyh-Shibg [27], fuzzy if-then rules are the cases where membership functions characterize the following statement if “ $A = x$ then $B = y$ ” where A and B are linguistic variables and x and y are linguistic values that are connected with membership functions.

A Sugino-type is also another kind of fuzzy if-then rules, where the premise part contains fuzzy sets only, whereas a nonfuzzy set defines the consequent part. It is also known as an adaptive neurofuzzy inference system (ANFIS) that is a type of fuzzy logic and neural network [27]. It has the ability to extract the advantages of both fuzzy logic and neural network in a single method [28]. It uses the algorithm of back-propagation and the least squared to learn the data to alter the membership function that assists the fuzzy to train the data to be modeled [29].

2.3. Functional Network (FN). A functional network (FN) is an extension of an ANN that comprises several layers of neurons linked to each other. Every processing neuron makes an explicit calculation: a scalar usually monotone f function of a weighted total of inputs. The f function, combined with the neurons, is constant, and the weights are learned from data utilizing some famous algorithms like the least-square fitting [30].

An FN comprises the input layer of input data, an output layer, single or many computing neuron layers that appraise

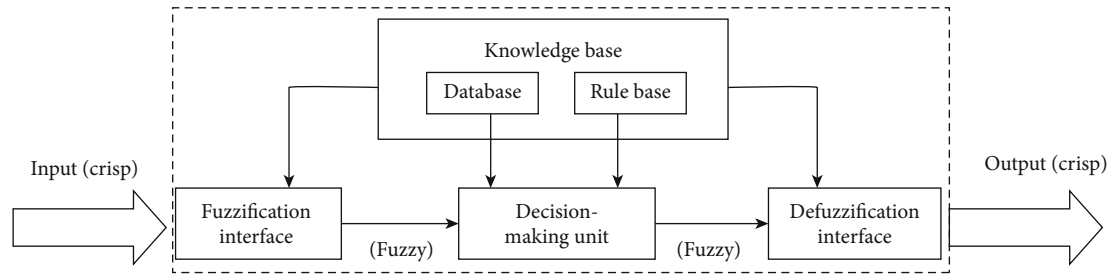


FIGURE 1: Parts of fuzzy logic.

a group of input values, comes from the input layer, and provides a group of output values to the output layer. The computing neuron layers are associated with each other, which mean that the output from one unit is able to work as a portion of the input to another neuron. When the input parameters are provided, the output is found by the type of functional network [31]. The differences between FN and ANN are that the weights in FN do not appear while the weights in ANN can appear. The neuron outputs of ANN are not the same while the neuron outputs of FN are coincident, and the neural functions in ANN are univariate, while in FN, they are multivariate.

Loss of return is influenced by various factors such as fluid properties, formation properties, and several known and unknown parameters. Therefore, it is arduous to predict. Therefore, many researchers applied the artificial intelligence to solve problems related to lost circulation such as Anifowose et al. [32], Castillo [33], Moazzeni et al. [34], Toreifi et al. [35], Efendiyev et al. [36], Far and Hosseini [37], Solomon et al. [38], Manshad et al. [39], Al-Hameedi et al. [40], Alkinani et al. [41], Abbas et al. [42], Cristofaro et al. [43], and Jahanbakhshi and Keshavarzi [44]. All these studies applied a single technique of AI to predict either the type of losses, the amount of losses, or the loss treatment, besides using many input parameters that are difficult to access in every well. None of these studies predicted the zones of the losses or used the real-time mechanical surface drilling parameters in their predictions.

The objective of this study is to predict the lost circulation zones using surface drilling parameters obtained by real-time drilling sensors. Three artificial intelligence techniques, namely, ANN, fuzzy logic (FL), and FN, are compared to achieve the objective.

3. Methodology

Figure 2 summarizes the processes of the methodology used in this study to predict the loss zones.

3.1. Data Acquisition. Three onshore wells were selected, where the lost circulation records and the mechanical surface drilling parameters were used for this study. The data were acquired on a per-foot basis from real-time sensors. The loss records include the flow out of the well (FLWOUT %) and the depth of the losses. The mechanical surface drilling parameters include the depth (D), hook height (HKHT), hook loud (HKL), flow pump (FPWPMP), rate of penetra-

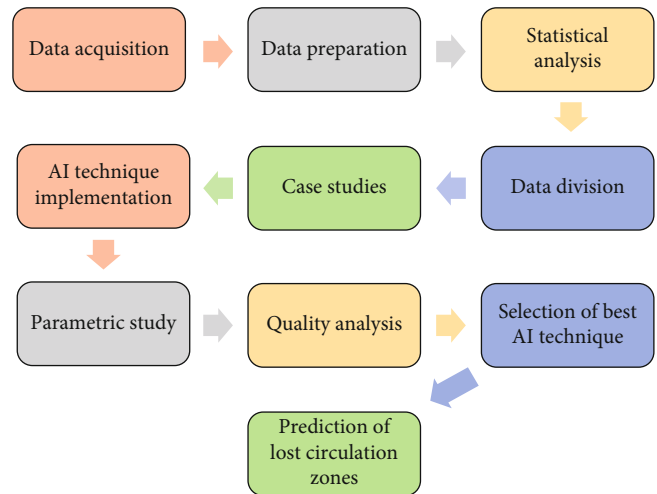


FIGURE 2: Steps of the methodology used.

tion (ROP), string rotary speed (RPM), standpipe pressure (SPP), drilling torque (TORQUE), and weight on bit (WOB). The circulation loss occurred in the three wells, and the drilling was continued until reaching the end of the section without curing the losses. Figure 3(a) shows the collected data of WOB versus the depth in Well A.

3.2. Data Preparation. Firstly, the data were collected from all operations involved in the three phases of the overall drilling process, i.e., drilling, tripping, and running the casing. All missing values, such as 999 values, and negative values, were removed. The second step was to include only the data in the drilling phase operation, while the data from the other phases were considered as unwanted. The data from the drilling phase operation were reorganized based on fresh footage, which requires human involvement to mark the minimum and maximum depths reached and eliminate any depth values beyond the maximum depth. Then, any footage values less than the previous were removed and will be considered a tripping operation. Figure 3(b) shows the data of WOB versus the depth from Well A after removing the random values and selecting the drilling phase operation.

The next stage was to further smoothen the data by eliminating the outliers or noise. Many filtration techniques, which have been implemented to allow data automation in the future, were applied to smooth the data. These techniques included *movmean*, *movmedian*, *Gaussian*, *lowess*, *loess*,

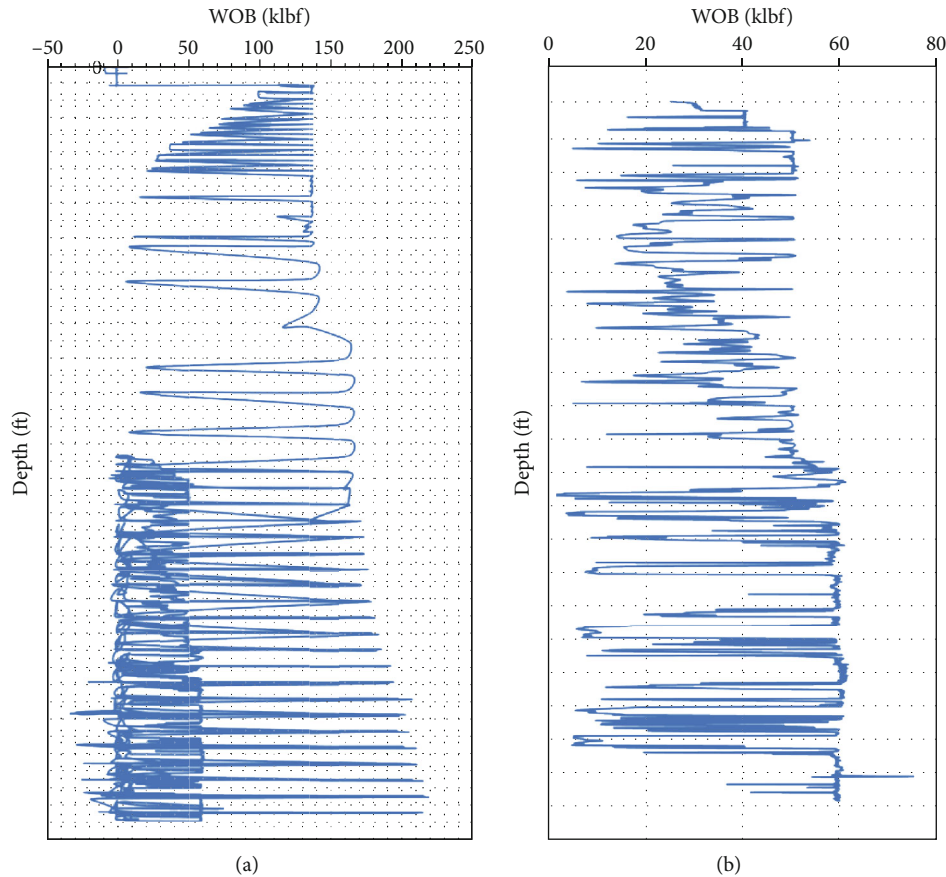


FIGURE 3: The collected WOB vs. depth from Well A (a) before filtration and (b) after filtration.

rloess, *rlloess*, and *sgolay*. Figure 4 shows the application of all the techniques on the WOB parameter from Well A.

The best filtration technique is *movmean*, which ensures that most of the data are preserved without significantly altering them. The performance of the *sgolay* filter was also found to be close to that of *movmean*; however, when processing big data in real-time, *movmean* is preferred as it requires less computing power [45]. The *movmean* technique was also applied to filter the WOB parameter from Well A with a span of 2, 4, 6, 8, 10, and 5 to determine the optimum noise reduction while retaining the data structure. The span of 5 is the best for data smoothing.

Regarding the output, the only action taken was to prepare the data in the proper format. As the two relevant conditions for each well section are losses or no losses, the data were arranged, as shown in Table 1, with the corresponding condition identified with 1 or 0.

3.3. Statistical Analysis. The best approach to examine the influence of different parameters on the loss of circulation is by performing a statistical analysis. Data diversity was assessed through a comprehensive statistical analysis. Statistical description contains a minimum, maximum, mean, range, mode, variation, kurtosis, skewness, and standard deviation. Table 2 shows the statistical analysis of Well A.

3.4. Data Division. Data from Well A were used to build the three AI models, while data from Well B and Well C were used to validate the AI models. The Well A data were randomly divided into two parts: the first part was used to train the model and the second part was used to test its ability to predict the values of the relevant parameters. The percentages of data used for training and testing were selected by trial and error.

Initial ANN runs were conducted using several different percentages of data for training and testing to select the best proportion on a trial and error basis. The previously identified six input parameters of FLWPMP, ROP, RPM, SPP, TORQUE, and WOB were used in these trials. Figure 5 shows the results of all the trials for the selected training and testing data distributions. The results reveal that the distribution of 75% training dataset and 25% testing dataset is the best in terms of both *R* and RMSE.

3.5. Case Studies. Several cases were evaluated to examine the impact of the input parameters on the prediction of lost circulation zones and enhance AI accuracy by removing the unnecessary parameters. In each run, the effect of a specific parameter on predicting the lost circulation zones was evaluated, while keeping the other parameters constant. Figure 6 presents the results of all the trials defined in Table 3 for

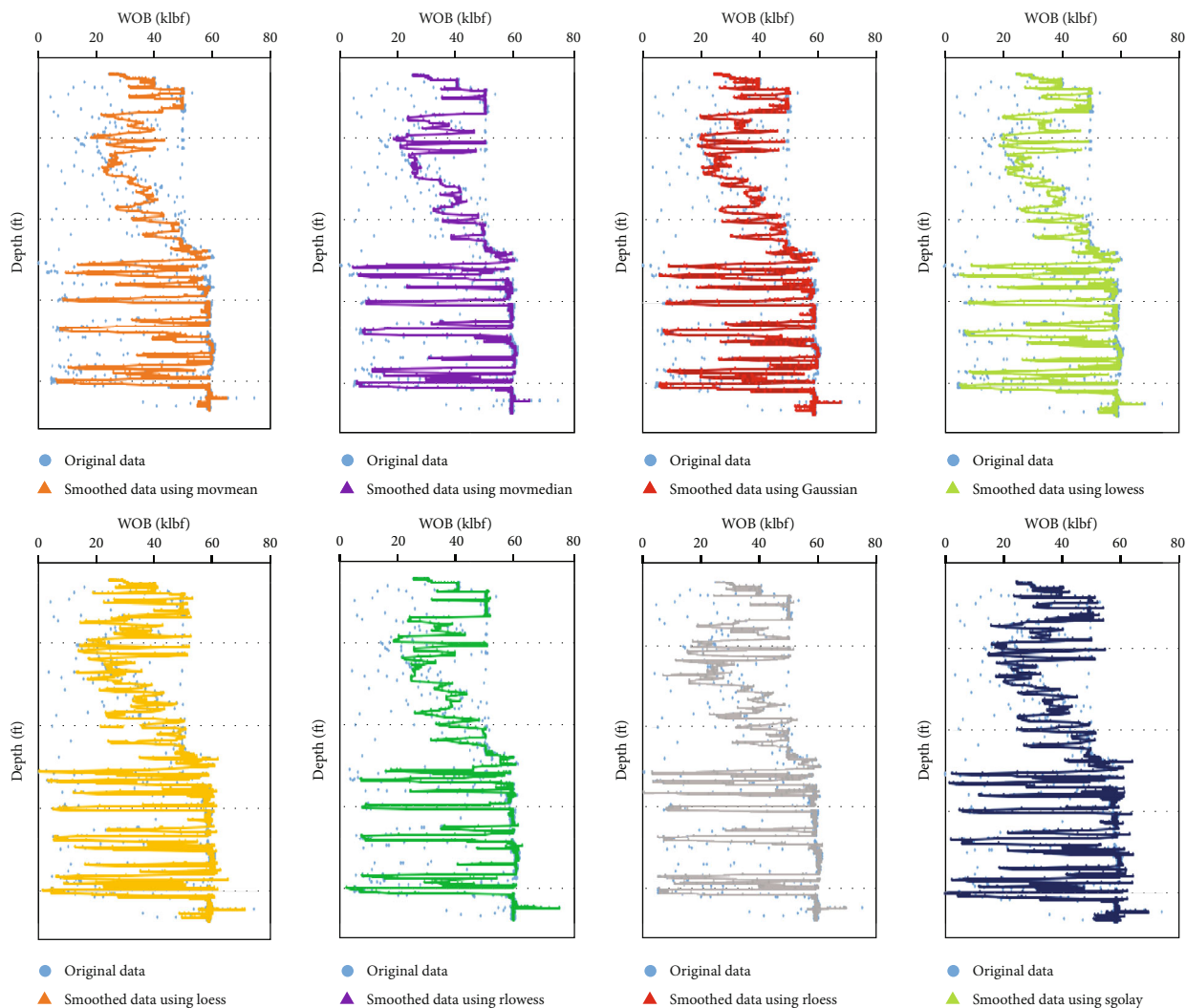


FIGURE 4: Filtration techniques applied to WOB from Well A.

TABLE 1: Input and output setup.

	Depth zone	FLWPMPS (gal/min)	Input				WOB (klbf)	Output	
			ROP (ft/h)	RPM (rpm)	SPP (psi)	TORQUE (klbf-ft)		Losses	No losses
Loss zone	1	1006	135.4	65.8	2451.5	11.9	48.9	0	1
	2	1006	167.7	66	2431.4	11.8	45.5	0	1
	3	970	162.1	64	2159.9	10.3	34.6	0	1
No loss zone	1	850	140.2	60.8	1520.5	7.5	14.7	1	0
	2	790	116.2	59.6	1225.8	6.1	8.9	1	0
	3	769	92.1	58.2	1097.1	5.7	8.2	1	0

selecting the input parameters. Based on the correlation coefficient and the root mean square error, the six input parameters (Trial 1) of FLWPMPS, ROP, RPM, SPP, TORQUE, and WOB were selected to predict lost circulation zones. Trials 1, 4, and 7 are giving the best results based on various combinations of independent parameters. Trial 1 requires six parameters that are one parameter more than each of

Trial 4 and Trial 7. Nevertheless, we believe that Trial 1 includes parameters that are physically important to detect the loss of circulation zones such as RPM and WOB. In contrast, Trials 4 and 7 are missing at least one of them.

The trials focused only on the mechanical surface drilling parameters that are very important in drilling and readily available for each well. All six parameters show significant

TABLE 2: Statistical analysis of all the parameters in Well A.

Parameter	FLWOUT (%)	HKHT (ft)	HKLI (klbf)	FLWPMPS (gal/min)	ROP (ft/h)	RPM (RPM)	SPP (psi)	TORQUE (kft-lbf)	WOB (klbf)
Minimum	0	0.218	160.946	721.000	9.260	49.720	1002.727	4.480	5.128
Maximum	54.480	88.323	249.116	1201.800	200.408	100.280	2753.440	14.344	68.372
Mean	9.664	44.615	191.665	1072.861	54.497	73.125	2193.087	8.686	49.335
Range	54.480	88.105	88.170	480.800	191.148	50.560	1750.713	9.864	63.244
Mode	0.000	63.043	200.476	1006.000	18.060	80.000	2459.280	8.696	58.956
Skewness	1.028	-0.051	0.749	-0.350	1.571	-0.034	-0.794	0.487	-1.305
Standard deviation	15.254	24.063	15.406	126.587	53.358	12.418	368.205	1.703	13.656

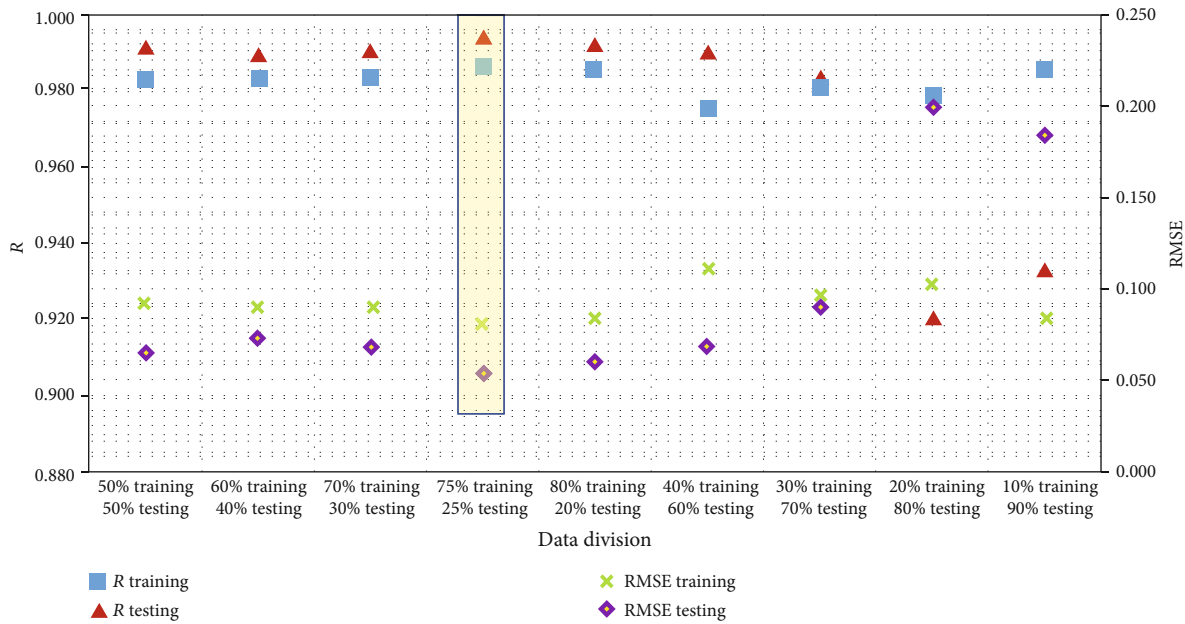


FIGURE 5: The results of the trials conducted for selecting the training and testing data distributions.

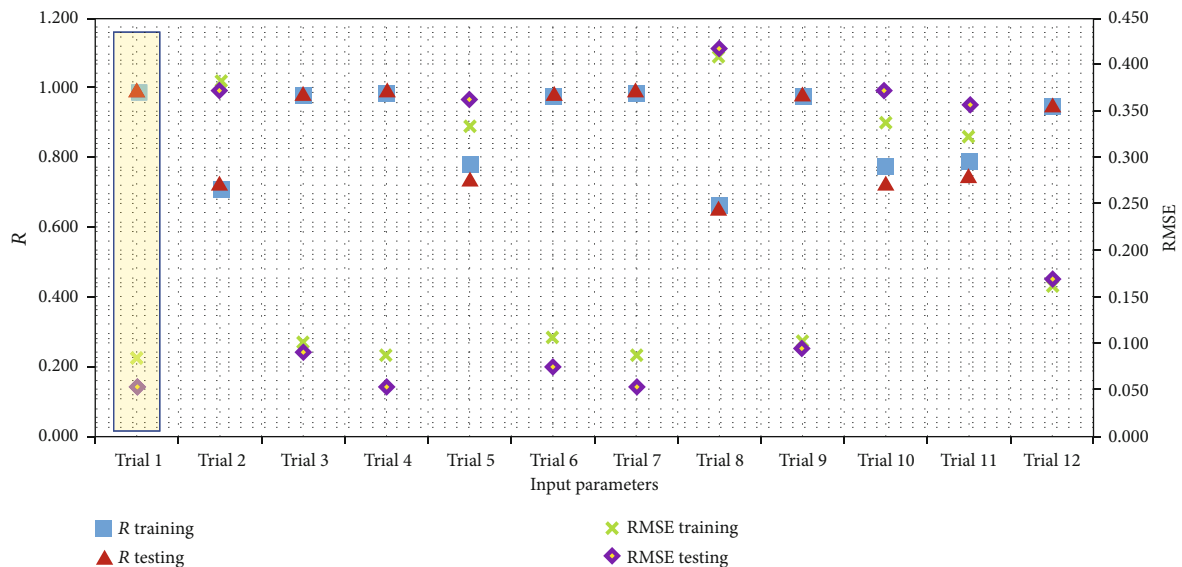


FIGURE 6: The results of the evaluation to select the important input parameters.

TABLE 3: Parameters used for the trials for selecting input parameters.

Trial #	Input parameters
Trial 1	FLWPMPS, ROP, RPM, SPP, TORQUE, WOB
Trial 2	ROP, RPM, SPP, TORQUE, WOB
Trial 3	FLWPMPS, RPM, SPP, TORQUE, WOB
Trial 4	FLWPMPS, ROP, SPP, TORQUE, WOB
Trial 5	FLWPMPS, ROP, RPM, TORQUE, WOB
Trial 6	FLWPMPS, ROP, RPM, SPP, WOB
Trial 7	FLWPMPS, ROP, RPM, SPP, TORQUE
Trial 8	RPM, SPP, TORQUE, WOB
Trial 9	FLWPMPS, SPP, TORQUE, WOB
Trial 10	FLWPMPS, ROP, TORQUE, WOB
Trial 11	FLWPMPS, ROP, RPM, WOB
Trial 12	FLWPMPS, ROP, RPM, SPP

changes in their values after the occurrence of lost circulation, as shown in Figures 7(a)–7(c). Hook loud was excluded from the trials as it did not change after the event of lost circulation and due to the availability of WOB, which is an alternative for hook loud.

3.6. Implementation of AI Techniques. After selecting the best input parameters for constructing the AI models, the next step is to apply artificial intelligence techniques. As previously stated, the artificial intelligence tools ANN, FL, and FN were used. MATLAB 2016 software was used to implement the AI methods.

3.6.1. ANN Implementation. Many trials were conducted to select the optimum number of neurons, training functions, transfer functions, and the network function using ANN to predict the zones of lost circulation. One layer with different numbers of neurons was used, and the results are shown in Figure 8. The results are very close to each other and are of high accuracy. Increasing the number of neurons will increase the computational time and will result in a large number of weights and biases, which in turn will increase the number of constants in the correlation equation. Therefore, five neurons are selected based on their higher accuracy in the testing part and to keep the network fast and efficient.

The performance of 13 training functions was evaluated to determine the optimum training function. The results presented in Figure 9 indicate that the optimum training function that has the highest accuracy is *Trainbr*. Thirteen transfer functions were also evaluated to observe their impacts on the prediction of lost circulation zones. Figure 10 summarizes the performance of these transfer functions and reveals that *logsig* is the optimum transfer function.

Then, ten network functions were assessed to determine their influence in the prediction of lost circulation zones. The results presented in Figure 11 show that the values are close to each other. Even though *Fitnet*, *Newff*, and *Newfit* produce the same accuracy in the prediction of the lost circu-

lation zones, the *Fitnet* network function is the one selected in the ANN model.

The ANN model was built with five neurons in one internal layer, and with a training function of *Trainbr*, the transfer function of *Logsig*, and network function of *Fitnet*.

3.6.2. FL Implementation. Two tools of fuzzy logic were performed to determine their impacts on the prediction of the lost circulation zones. The first tool is Mamdani FIS (*Genfis 1*), and the second tool is Sugeno FIS (*Genfis 2*). Mamdani FIS is not suitable because it has a long processing time. Many trials were performed using *Genfis 2* (Sugeno FIS) to determine the best selection of the epoch size and the radius. The results in Figure 12 indicate that the prediction accuracy does not change after an epoch size of 70, and based on the highest correlation coefficient and the lowest root mean square error, an epoch size of 70 was selected.

Several radii were evaluated at 70 iterations, and the results are shown in Figure 13. A radius of 0.5 produced the highest accuracy. Therefore, the fuzzy logic model was built using Sugeno FIS with an epoch size of 70 and a radius of 0.5.

3.6.3. FN Implementation. Several FN trials were run to select the best methods and types of functional networks to predict the zones of lost circulation. Many procedures were evaluated for both linear and nonlinear types to find their effect on the prediction of the lost circulation zones. These procedures include functional network backward-forward method (FNBFM), functional network forward-backward method (FNFBM), functional network backward-exhaustive method (FNBEM), functional network forward-selection method (FNFSM), and functional network exhaustive selection method (FNESM). The results are shown in Figure 14. Based on the highest R and the lowest RMSE, the Type 3 of FNFBM has the highest accuracy. Therefore, the FN model, with the features of functional network forward-backward method (FNFBM) and nonlinear Type 3 was selected to predict the lost circulation zones.

4. Results and Discussion

The ANN implementation showed that an ANN, of five neurons in one internal layer, with the training function of *Trainbr*, the transfer function of *Logsig*, and the network function of *Fitnet*, gives the best performance with the highest correlation coefficient and the lowest root mean square error. The ANN predicted the lost circulation zones in the training part (75% of the data) with R of 0.987 and RMSE of 0.081, as shown in Figure 15(a). In the testing part (25% of the data), the ANN predicted the lost circulation zones with R of 0.994 and RMSE of 0.053, as indicated in Figure 16(a).

The results of the prediction are also presented in a confusion matrix. The results for the training set, depicted in Figure 17(a), show that the ANN model was able to predict 735 out of 742 locations of the lost circulation zones correctly, i.e., 99.1%, and the ANN model was not able to accurately predict only 7 locations, representing 0.9%. In the zones where losses do not occur, the ANN model was able

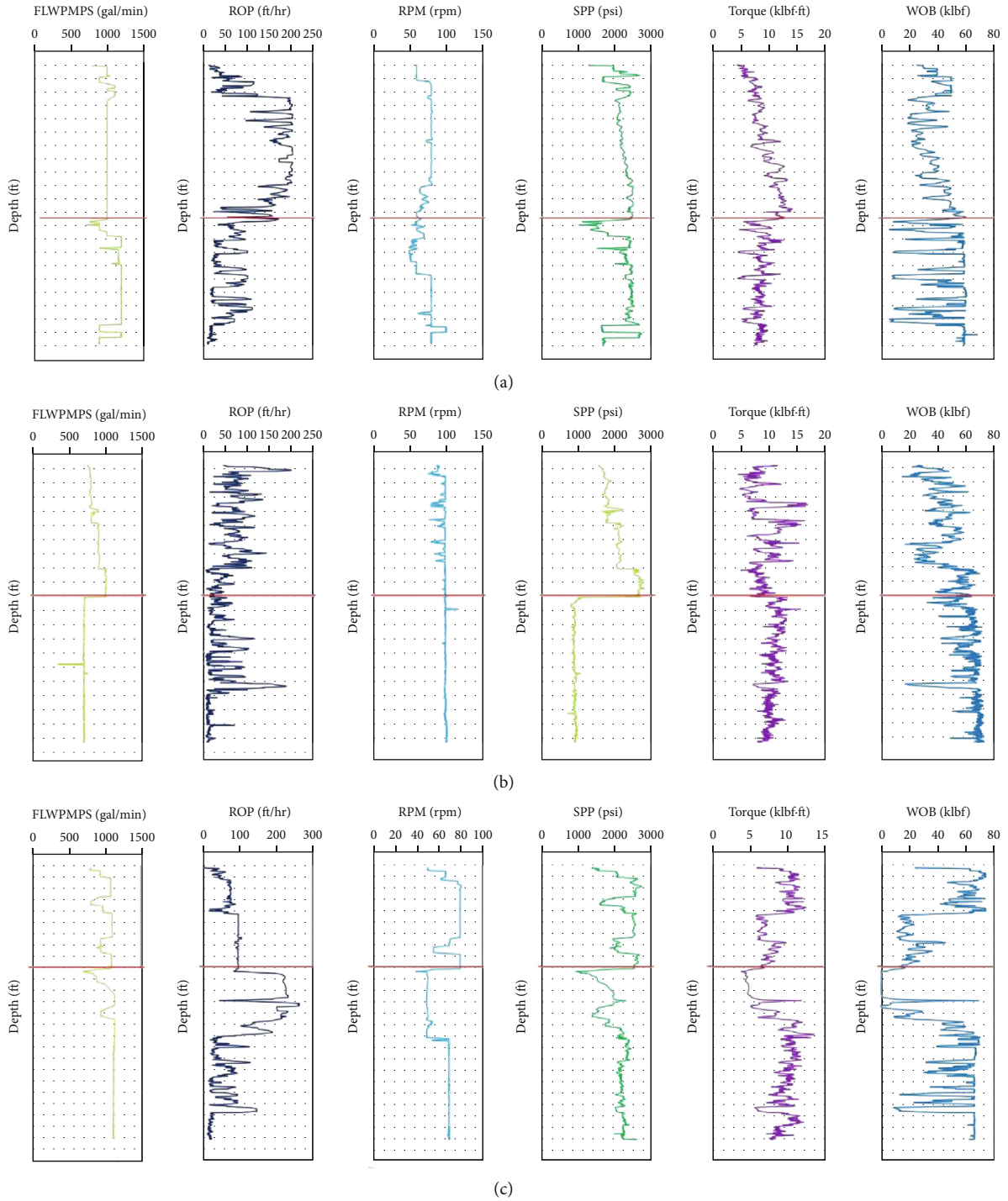


FIGURE 7: (a) The change of the six surface parameters after the occurrence of lost circulation in Well A. (b) The change of the six surface parameters after the occurrence of lost circulation in Well B. (c) The change of the six surface parameters after the occurrence of lost circulation in Well C.

to predict all 321 locations correctly. Considering all the locations of lost circulation in the two zones, the ANN model correctly predicted 99.3% of the locations, and the prediction of only 0.7% of the locations are incorrect. The results for the testing dataset shown in Figure 18(a) reveal that the ANN

model was able to predict 263 out of 264 locations in the lost circulation zones correctly, i.e., 99.6% of the locations, and it could not precisely predict only 1 point, representing 0.4% of the locations. In the zones where losses do not occur, the ANN model was able to predict all 90 locations correctly.

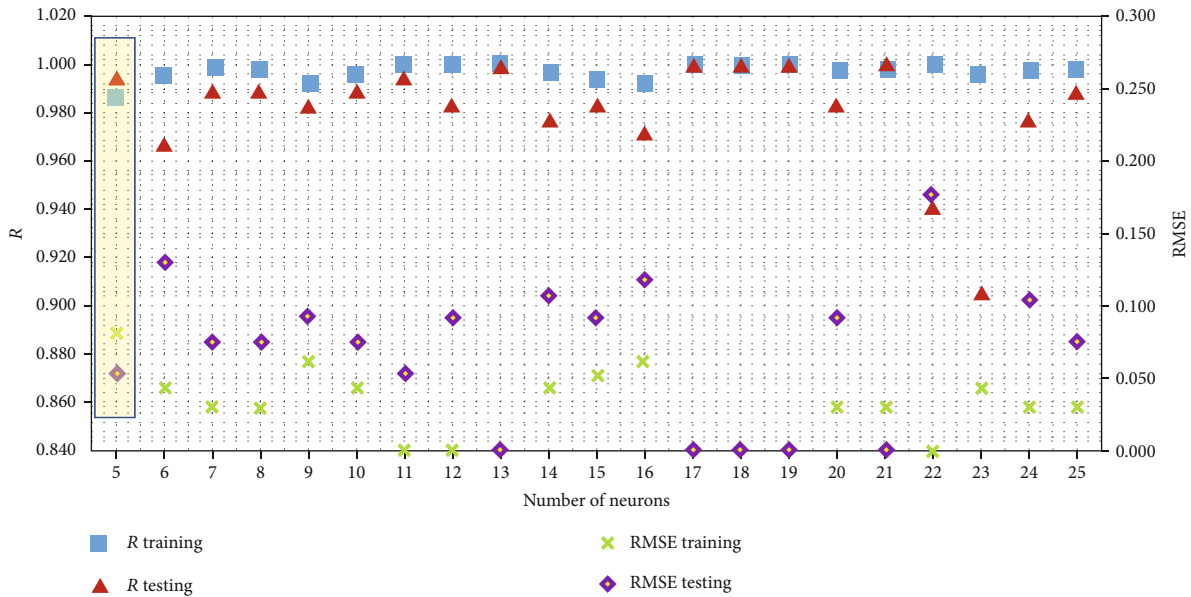


FIGURE 8: Impact of the number of neurons in one layer of the training and testing datasets.

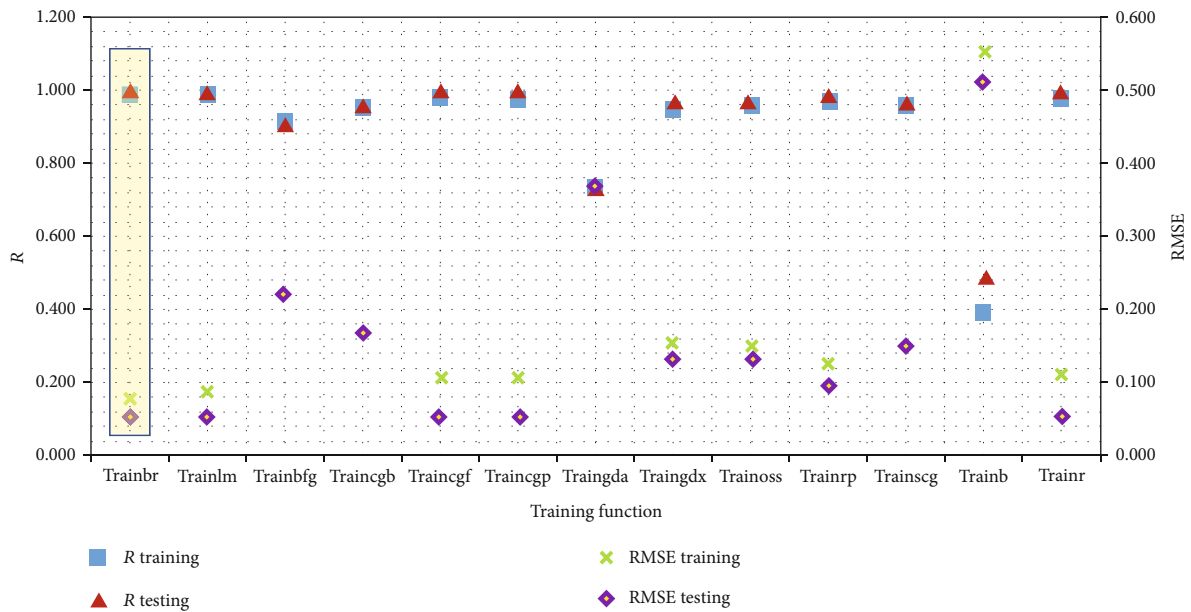


FIGURE 9: Results of using different training functions for the training and testing datasets.

Considering all the locations in the two zones, the ANN model correctly predicted 99.7% of the locations, and the prediction of only 0.3% of the locations is incorrect.

The FL implementation showed that the best performance of the FL that produces the highest correlation coefficient and the lowest root mean square error is with the features of Sugeno FIS, an epoch size of 70, and a radius of 0.5. The FL model predicted the lost circulation zones in the training part (75% of the data) with R of 0.993 and RMSE of 0.053, as shown in Figure 15(b). In the testing part (25% of the data), the FL predicted the lost circulation zones with R of 0.993 and RMSE of 0.053, as indicated in Figure 16(b).

A confusion matrix also presents the results of the prediction. The results for the training dataset shown in Figure 17(b) indicate that the FL model was able to predict 739 out of 742 locations in the lost circulation zones correctly, with an accuracy of 99.6%. In comparison, it mispredicted only 3 locations, representing 0.4% of the data. In the zones where losses do not occur, the FL model was able to predict all 321 locations correctly. Considering all the locations in the two zones, FL correctly predicted 99.7% of the locations, with only 0.3% mispredicted. The results for the testing dataset depicted in Figure 18(b) show that the FL model was able to predict 263 out of 264 locations in the lost circulation zones correctly, with an accuracy of 99.6%. In

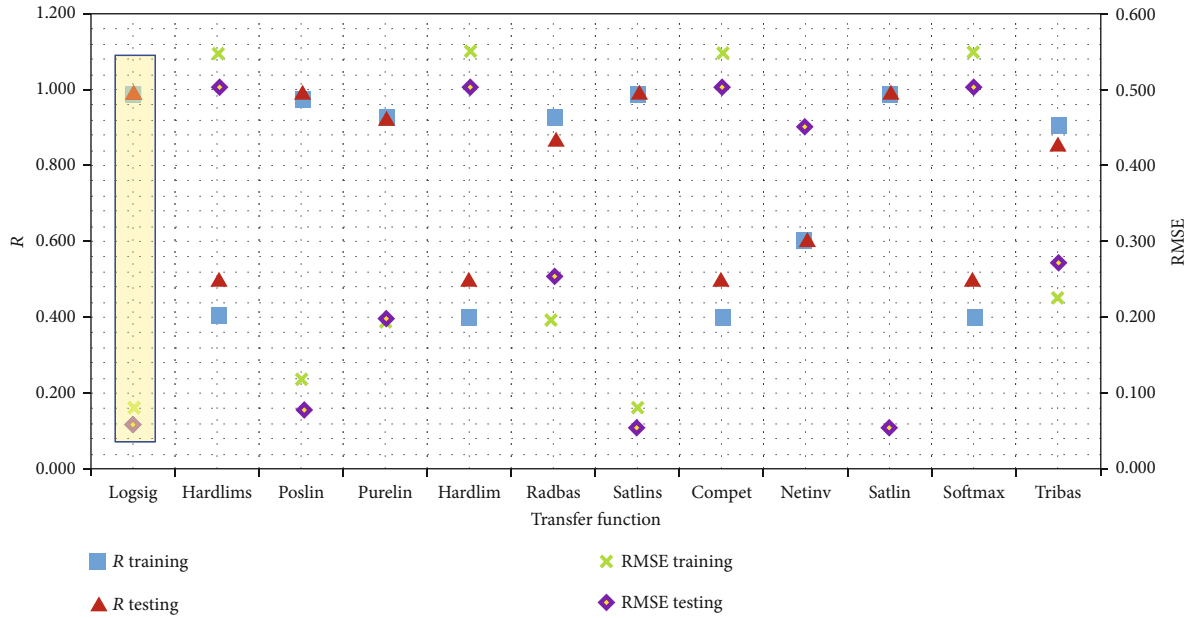


FIGURE 10: Influence of using different transfer functions for the training and testing datasets.

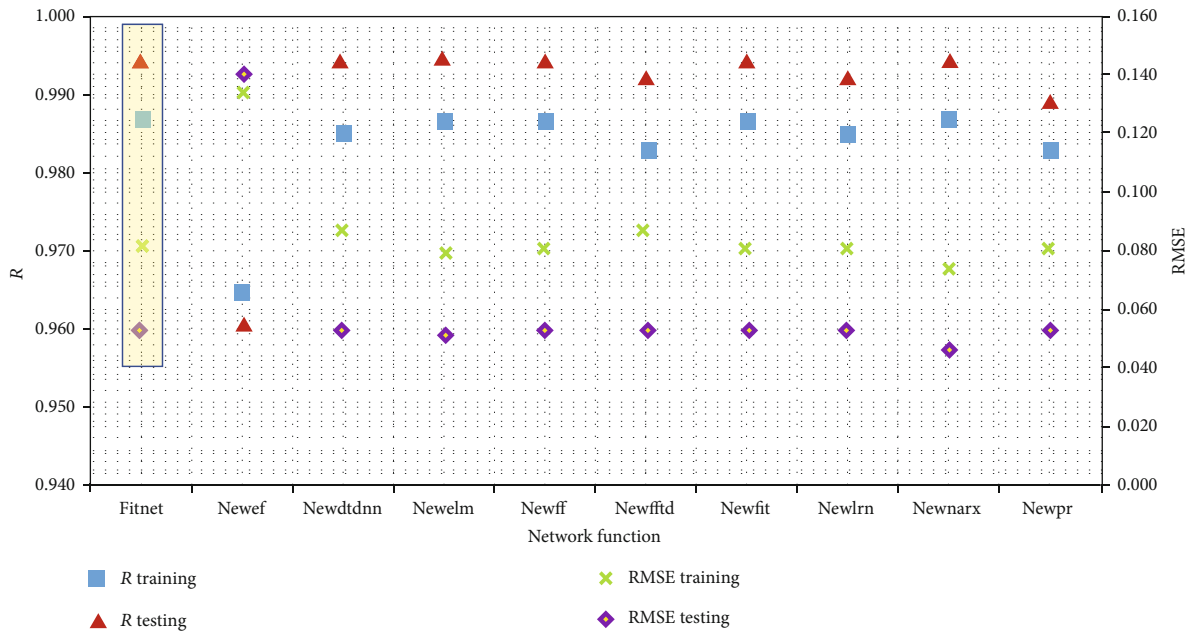


FIGURE 11: Effects of using different network functions for the training and testing datasets.

contrast, it mispredicted only one location, representing 0.3% of the data. In the zones where losses do not occur, the FL model was able to predict all 90 locations correctly. Considering all the locations in the two zones, the FL correctly predicted 99.7% of the locations, with only 0.3% mispredicted.

The FN implementation showed that the best performance of the FN that produces the highest correlation coefficient and the lowest root mean square error is with the features of FNFBM and nonlinear Type 3. The FN model predicted the lost circulation zones in the training part (75% of the data) with R of 1 and RMSE of 0, as shown in

Figure 15(c). In the testing part (25% of the data), the FN model predicted the lost circulation zones with R of 0.985 and RMSE of 0.075, as indicated in Figure 16(c).

A confusion matrix also presents the results of the prediction. The results for the training dataset depicted in Figure 17(c) indicate that the FN model was able to predict all 742 locations in the lost circulation zones correctly. In the zones where losses do not occur, the FN model was also able to predict all 321 locations correctly. Considering all the locations in the two zones, FN correctly predicted 100% of the data locations. The results for the testing dataset shown

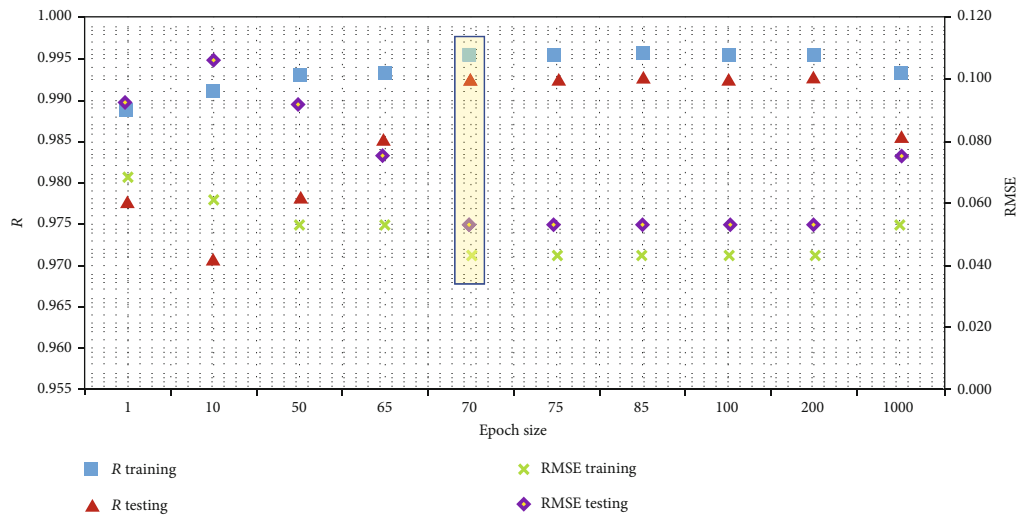


FIGURE 12: Influence of the iteration number in Sugeno FIS for the training and testing datasets.

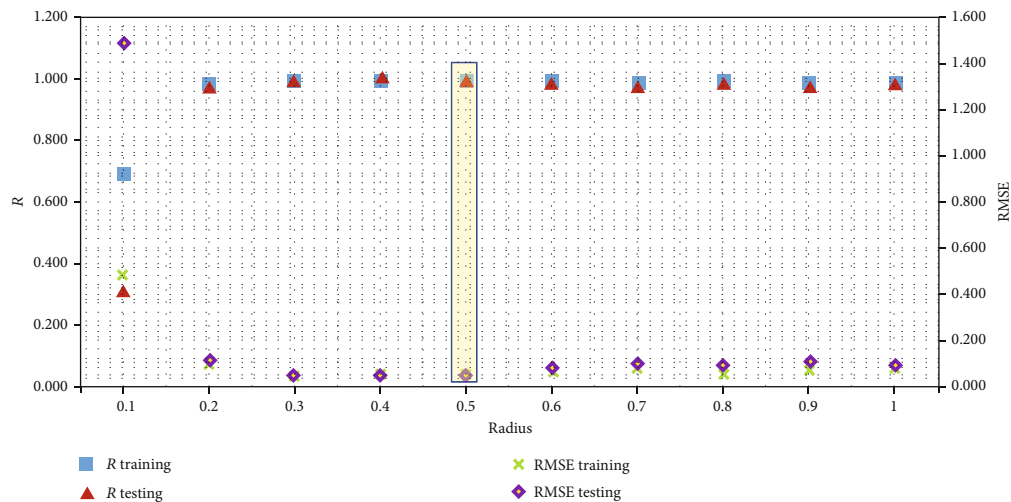


FIGURE 13: Impact of the radius in Sugeno FIS for the training and testing datasets.

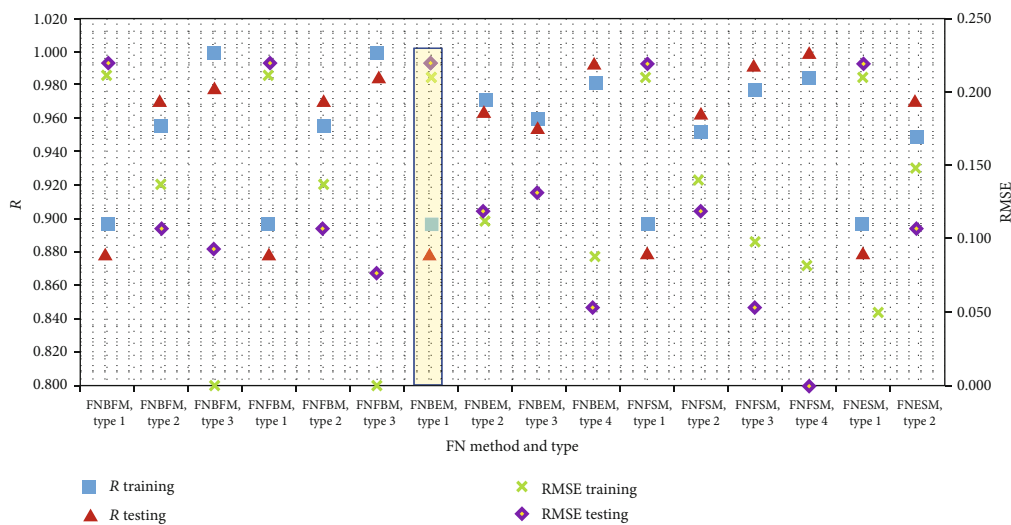


FIGURE 14: Influence of the FN methods and types for the training and testing datasets.

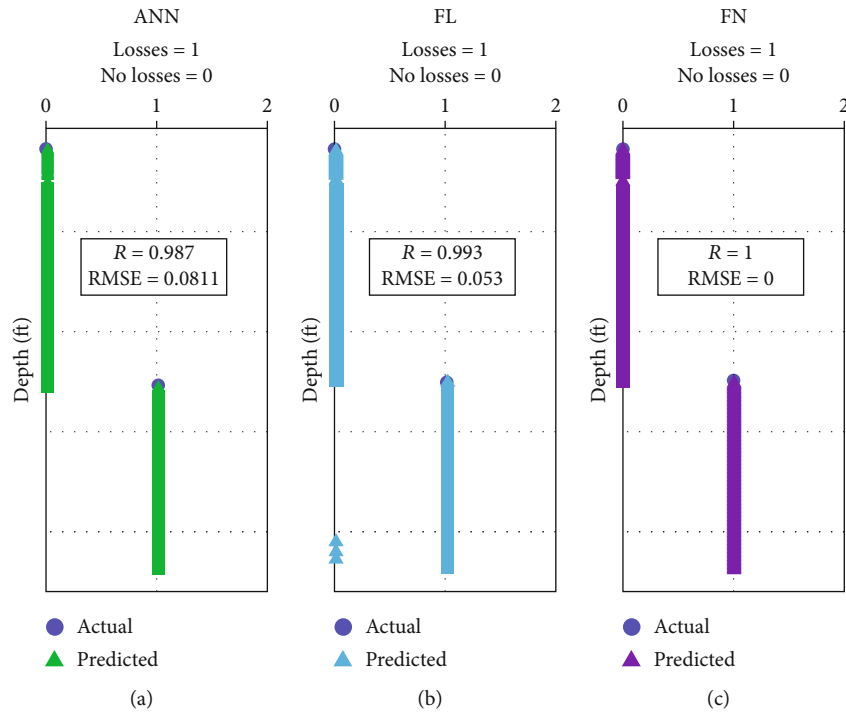


FIGURE 15: Prediction of the lost circulation zones using the five AI models for the training dataset.

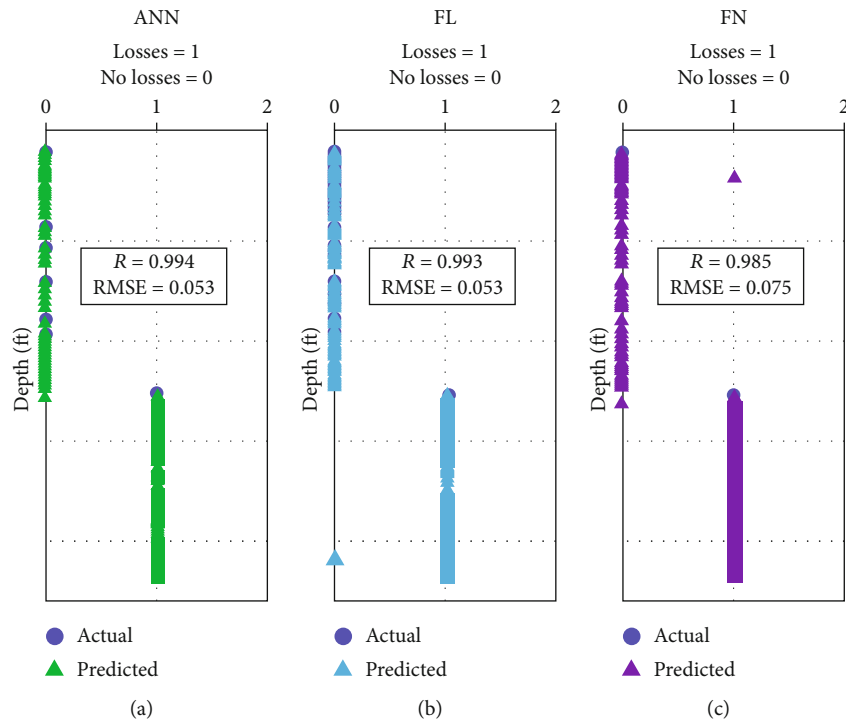


FIGURE 16: Prediction of the lost circulation zones using the five AI models for the testing dataset.

in Figure 18(c) reveal that the FN model was able to predict 263 out of 264 locations in the lost circulation zones correctly, with an accuracy of 99.6%. In the zones where losses

do not occur, the FN model was able to predict 89 locations out of 90 locations correctly, representing 98.8% of the data, and it mispredicted only one location, representing 1.1% of

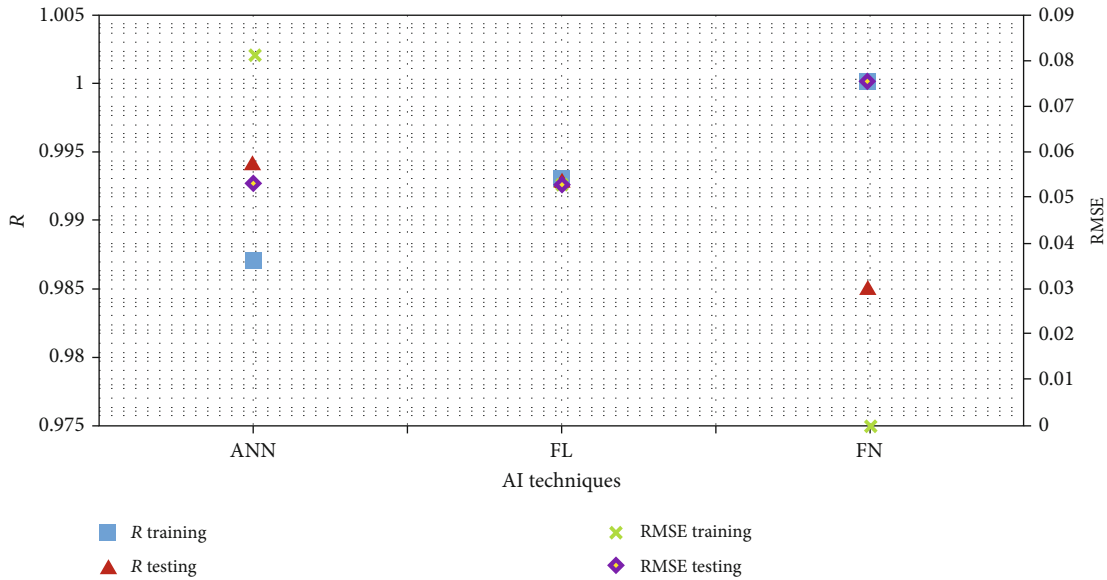


FIGURE 19: Comparison of the five AI techniques for the training and testing datasets.

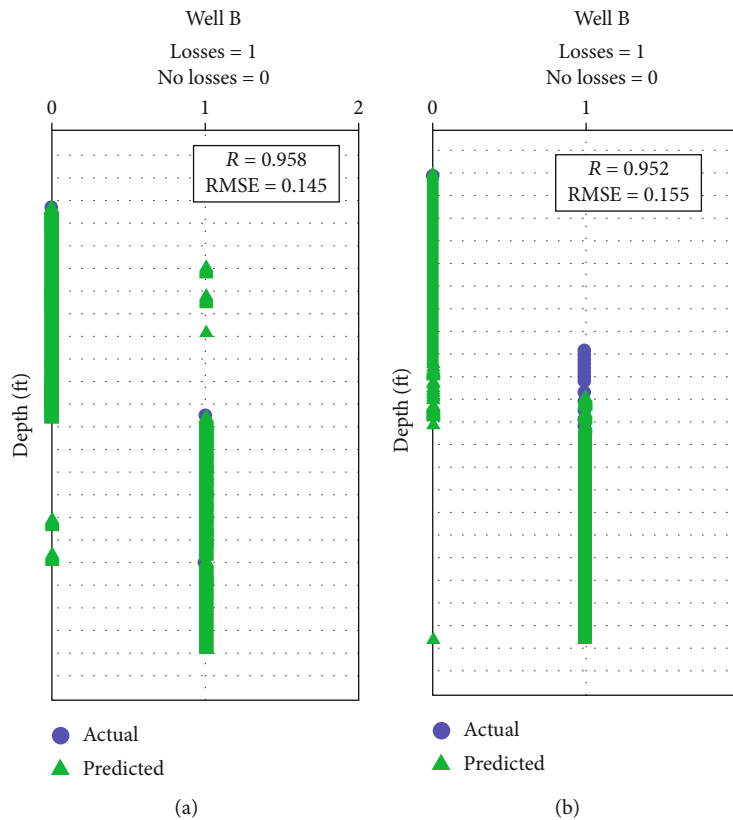


FIGURE 20: Prediction of the lost circulation zones using the ANN model in Well B (a) and Well C (b).

was selected for the validation using the data from Well B and Well C because of its highest accuracy in the testing part.

In Well B, ANN was able to predict the lost circulation zones with high accuracy ($R = 0.958$ and $RMSE = 0.145$), and the results of the evaluation are shown in Figure 20(a). The confusion matrix shown in Figure 21(a) indicates that the ANN model was able to predict 2067 out of 2103 loca-

tions in the lost circulation zones correctly, with an accuracy of 98.3%, with only 36 locations, representing 1.7%, not predicted correctly. In the zones where losses do not occur, the ANN model was able to predict 745 out of 769 locations correctly, with an accuracy of 96.9%, while only one location, representing 3.1% of the data, was mispredicted. Considering all the locations in the two zones, the ANN model correctly



FIGURE 21: Confusion matrix for the prediction of lost circulation zones in Well B (a) and Well C (b) using the ANN technique.

predicted 97.9% of the locations, while only 2.1% was predicted correctly.

In Well C, the ANN model was able to predict the lost circulation zones with high accuracy ($R = 0.952$ and $RMSE = 0.155$), and the results of the evaluation are shown in Figure 20(b). The confusion matrix shown in Figure 21(b) indicates that the ANN model was able to predict 863 out of 894 locations in the lost circulation zones correctly, with an accuracy of 96.5%, with only 31 locations, representing 3.5% of the data, not predicted correctly. In the zones where losses do not occur, the ANN model was able to predict all 400 locations correctly. However, considering the two zones, the ANN technique correctly predicted 97.6% of the data, while only 2.4% were mispredicted.

7. Conclusions

This study evaluated three AI techniques to predict the lost circulation zones based only on six mechanical surface drilling parameters. These techniques are functional networks (FN), artificial neural networks (ANN) and fuzzy logic (FL). The six parameters are real-time measurements of flow pump (FPWPMP), rate of penetration (ROP), string rotary speed (RPM), standpipe pressure (SPP), drilling torque (TORQUE), and weight on bit (WOB). More than 4500 real-field data points from three wells were used in the evaluation. The following conclusions are drawn from the results of this study:

- (i) The AI techniques were trained and tested using the data from Well A to predict the lost circulation zones with high precision
- (ii) The accuracy of the ANN was $R = 0.987$ and $RMSE = 0.0811$ for training and $R = 0.994$ and $RMSE = 0.053$ for testing
- (iii) The accuracy of the FL was $R = 0.993$ and $RMSE = 0.0531$ for training and $R = 0.993$ and $RMSE = 0.053$ for testing
- (iv) The accuracy of the FN was $R = 1$ and $RMSE = 0$ for training and $R = 0.985$ and $RMSE = 0.0752$ for testing
- (v) The ANN was the best technique due to its highest accuracy in the testing. So, its model was validated using data from the second and third wells (Well B and Well C, respectively), which are unseen
- (vi) The ANN was able to identify the lost circulation zones in Well B and Well C with a high performance of $R = 0.958$ and $RMSE = 0.145$ and $R = 0.952$ and $RMSE = 0.155$, respectively
- (vii) The main advantage of AI techniques is their simplicity, which allows the prediction of lost circulation zones from only the mechanical surface drilling parameters that are readily available in each well

The future direction of this work is to validate the developed ANN model on several wells and improve the ANN model to be able to predict the amount of the losses using the real-time surface drilling parameters, where the mud properties, formation properties, and fracture's length and width are also major parameters that need to be considered to predict the loss amount.

Nomenclature

AI:	Artificial intelligence
FN:	Functional network
ANN:	Artificial neural networks
FL:	Fuzzy logic
Trainbr:	Bayesian regularization
Trainlm:	Levenberg-Marquardt backpropagation
Trainbfg:	BFGS quasi-Newton backpropagation
Traincgb:	Powell-Beale conjugate gradient backpropagation
Traincgb:	Powell-Beale conjugate gradient backpropagation
Traincgf:	Fletcher-Powell conjugate gradient backpropagation
Traincgp:	Polak-Ribière conjugate gradient backpropagation
Trainгда:	Gradient descent with adaptive learning rule backpropagation
Trainгдаx:	Gradient descent with momentum and adaptive learning rule backpropagation
Trainoss:	One-step secant backpropagation
Trainr:	Random-order incremental training with learning functions
Trainrp:	Resilient backpropagation (Rprop)
Trainscg:	Scaled conjugate gradient backpropagation
Compet:	Competitive transfer function
Hardlim:	Hard-limit transfer function
Hardlims:	Symmetric hard-limit transfer function
Logsig:	Log-sigmoid transfer function
Netinv:	Inverse transfer function
Poslin:	Positive linear transfer function
Purelin:	Linear transfer function
Radbas:	Radial basis transfer function
Satlin:	Saturating linear transfer function
Satlins:	Symmetric saturating linear transfer function
Softmax:	Softmax transfer function
Tansig:	Hyperbolic tangent sigmoid transfer function
Tribas:	Triangular basis transfer function
Newcwf:	Create cascade-forward backpropagation network
Newtdnn:	Create distributed time delay neural network
Newelm:	Create Elman backpropagation network
Newff:	Create feedforward backpropagation network
Newfftd:	Create feedforward input-delay backpropagation network
Newfit:	Create fitting network
Newlrn:	Create layered-recurrent network
Newnarx:	Feedforward backpropagation network with feedback from output to input
Newpr:	Create pattern recognition network
Fitnet:	Function fitting neural network
Genfis 1:	Mamdani FIS
Genfis 2:	Sugeno FIS
FNFBM:	Functional network forward-backward method
FNBEM:	Functional network backward-exhaustive method
FNFSM:	Functional network forward-selection method
FNESM:	Functional network exhaustive selection method
FLWPMPS:	Flow rate of the pump (gal/min)

ROP:	Rate of penetration (ft/hr)
SPP:	Stand pipe pressure (psi)
WOB:	Weight on bit (klbs)
RPM:	Rotary speed (rpm)
TORQUE:	Drilling torque (klbf)
R:	Correlation coefficient
RMSE:	Root mean square error

Data Availability

All the data required are included in the text.

Conflicts of Interest

The author declares no conflict of interest.

References

- [1] Halliburton, "High pressure high temperature (HP/HT) (n.d.)," 2011, <https://www.halliburton.com/en-US/ps/solutions/high-pressure-temperature/about-hp-ht-ht.html>.
- [2] M. Overholt, "Advantages of offshore oil rigs and drilling," 2017, <https://www.tigergeneral.com/advantages-offshore-oil-rigs-drilling/>.
- [3] A. Shadravan and M. Amani, "HPHT 101: what every engineer or geoscientist should know about high pressure high temperature wells," 2012, <https://www.onepetro.org/conference-paper/SPE-163376-MS>.
- [4] C. R. Miranda, J. L. Oliveira, G. M. S. Cavalcante et al., "Materials for controlling severe lost circulation-laboratory evaluation," in *SPE-185582-MS was presented at SPE Latin America and Caribbean Petroleum Engineering Conference*, Buenos Aires, Argentina, 2017.
- [5] J. Ramasamy and M. Amanullah, "Novel fibrous lost circulation materials derived from deceased date tree waste," in *SPE-187989-MS was presented at SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*, Dammam, Saudi Arabia, 2017.
- [6] S. P. Almagro, C. Frates, J. Garand, and A. Meyer, "Sealing fractures: advances in lost circulation control treatments," *Oil-field Review*, vol. 26, no. 3, pp. 4–13, 2014.
- [7] A. T. T. Al-Hameedi, H. H. Alkinani, S. Dunn-Norman et al., "Using machine learning to predict lost circulation in the Rumaila field, Iraq," in *SPE-191933-MS was presented at SPE Asia Pacific Oil and Gas Conference and Exhibition*, Brisbane, Australia, 2018.
- [8] J. Abdollahi, I. M. Carlsen, S. Mjaaland, P. Skalle, A. Rafiei, and S. Zarei, "Underbalanced drilling as a tool for optimized drilling and completion contingency in fractured carbonate reservoirs," in *SPE-91579-MS was presented at SPE/IADC Underbalanced Arabian Journal for Science and Engineering 1 3 Technology Conference and Exhibition*, Houston, Texas, 2004.
- [9] G. V. Chilingarian and P. Vorabutr, *Drilling and drilling fluids*, Elsevier, Amsterdam, Netherlands, 1981.
- [10] D. E. Caughron, D. K. Renfrow, J. R. Bruton et al., "Unique crosslinking pill in tandem with fracture prediction model cures circulation losses in deepwater Gulf of Mexico," in *SPE-74518-MS was presented at IADC/SPE Drilling Conference*, Dallas, Texas, 2002.

- [11] A. Lavrov, *Lost Circulation Mechanisms and Solutions*, Gulf Professional Publishing, Amsterdam, Netherlands, 2016.
- [12] R. Jadhav and S. Patil, "Acid-soluble thixotropic cement system for lost circulation challenges," in *SPE-193168-MS was presented at Abu Dhabi International Petroleum Exhibition & Conference*, Abu Dhabi, UAE, 2018.
- [13] T. R. Bratton, I. M. Rezmer-Cooper, J. Desroches, Y.-E. Gille, Q. Li, and M. McFayden, "How to diagnose drilling induced fractures in wells drilled with oil-based muds with real-time resistivity and pressure measurements," in *SPE-67742-MS was presented at SPE/IADC Drilling Conference*, Amsterdam, Netherlands, 2001.
- [14] J. U. Messenger, *Lost Circulation*, PennWell Publishing Company, 1981.
- [15] S. Mohaghegh, R. Arefi, S. Ameri, and D. Rose, "Design and development of an artificial neural network for estimation of formation permeability," *SPE Computer Applications*, vol. 7, 1995.
- [16] H. I. Bilgesu, U. Altmis, S. Ameri, S. Mohaghegh, and K. Aminian, "A new approach to predict bit life based on tooth or bearing failures," in *SPE Eastern Regional Meeting*, Pittsburgh, Pennsylvania, 1998.
- [17] K. Al-Azani, S. Elkatatny, A. Ali, E. Ramadan, and A. Abdurraheem, "Cutting concentration prediction in horizontal and deviated wells using artificial intelligence techniques," *Journal of Petroleum Exploration and Production Technology*, vol. 9, no. 4, pp. 2769–2779, 2019.
- [18] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE ASSP Magazine*, vol. 4, no. 2, pp. 4–22, 1987.
- [19] A. Ahmed, A. Ali, S. Elkatatny, and A. Abdurraheem, "New artificial neural networks model for predicting rate of penetration in deep shale formation," *Sustainability*, vol. 11, no. 22, p. 6527, 2019.
- [20] R. Schalkoff, *Artificial Neural Networks*, McGraw-Hill Education (India), The University of Michigan, 1997.
- [21] A. Ahmed, S. Elkatatny, A. Ali, M. Mahmoud, and A. Abdurraheem, "New model for pore pressure prediction while drilling using artificial neural networks," *Arabian Journal for Science and Engineering*, vol. 44, pp. 6079–6088, 2019.
- [22] S. Alarifi, S. AlNuaimi, and A. Abdurraheem, "Productivity index prediction for oil horizontal wells using different artificial intelligence techniques," in *SPE Middle East Oil & Gas Show and Conference*, Manama, Bahrain, 2015.
- [23] B. Kosko, *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice-Hall Inc., 1992.
- [24] Mathwork, *Fuzzy Logic Toolbox User's Guide*, MathWorks Inc., Natick, MA, USA, 2010, <http://citeseerx.ist.psu.edu/viewdoc/download?>
- [25] S. J. Cuddy and T. W. Putnam, "Litho-Facies and permeability prediction from electrical logs using fuzzy logic," *SPE Reservoir Evaluation & Engineering*, vol. 3, 1998.
- [26] R. B. C. C. Gharbi and G. A. Mansoori, "An introduction to artificial intelligence applications in petroleum exploration and production," vol. 49, 2005 *Journal of Petroleum Science and Engineering*, 2005.
- [27] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.
- [28] P. Tahmasebi and A. Hezarkhani, "A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation," *Computers & Geosciences*, vol. 42, pp. 18–27, 2012.
- [29] A. A. Anifowose and A. Abdurraheem, "Fuzzy logic-driven and SVM-driven hybrid computational intelligence models applied to oil and gas reservoir characterization," *Journal of Natural Gas Science and Engineering*, vol. 3, no. 3, pp. 505–517, 2011.
- [30] S. M. Elkatatny, T. Zeeshan, M. Mahmoud, A. Abdulazez, and I. M. Mohamed, "Application of artificial intelligent techniques to determine sonic time from well logs," in *50th U.S. Rock Mechanics/Geomechanics Symposium*, Houston, Texas, 2016 <https://www.onepetro.org/conference-paper/ARMA-2016-755>.
- [31] A. Ahmed, S. Elkatatny, A. Abdurraheem, M. Mahmoud, A. Z. Ali, and I. M. Mohamed, "Pore pressure prediction while drilling using fuzzy logic," in *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*, Dammam, Saudi Arabia, 2018.
- [32] F. Anifowose, T. Helmy, and A. Abdurraheem, "(PDF) Functional networks as a best-model selector in computational intelligence hybrid models for petroleum reservoir characterization," 2011, https://www.researchgate.net/publication/215662101_Functional_Networks_as_a_Best-Model_Selector_in_Computational_Intelligence_Hybrid_Models_for_Petroleum_Reservoir_Characterization.
- [33] E. Castillo, "Functional networks," *Neural Processing Letters*, vol. 7, no. 3, pp. 151–159, 1998.
- [34] A. R. Moazzeni, M. Nabaei, and S. G. Jegarluei, "Prediction of lost circulation using virtual intelligence in one of Iranian oil-fields," in *SPE-136992-MS was presented at Nigeria Annual International Conference and Exhibition*, Tinapa - Calabar, Nigeria, 2010.
- [35] H. Toreifi, H. Rostami, and A. K. Manshad, "New method for prediction and solving the problem of drilling fluid loss using modular neural network and particle swarm optimization algorithm," *Journal of Petroleum Exploration and Production Technology*, vol. 4, no. 4, pp. 371–379, 2014.
- [36] G. Efendiyev, P. Mammadov, I. Piriverdiyev, and V. Mammadov, "Clustering of geological objects using FCM-algorithm and evaluation of the rate of lost circulation," *Procedia Computer Science*, vol. 102, pp. 159–162, 2016.
- [37] P. Behnoud far and P. Hosseini, "Estimation of lost circulation amount occurs during under balanced drilling using drilling data and neural network," *Egyptian Journal of Petroleum*, vol. 26, no. 3, pp. 627–634, 2017.
- [38] O. Solomon, D. Adewale, and C. Anyanwu, "Fracture width prediction and loss prevention material sizing in depleted formations using artificial intelligence," in *SPE-189068-MS was presented at SPE Nigeria Annual International Conference and Exhibition*, 31, Lagos, Nigeria, 2017.
- [39] A. Manshad, H. Rostami, H. Niknafs, and A. Mohammadi, *In book: Heavy Oil*, pp. 243–253, Nova Science Publishers, Inc., NY, USA, 2017.
- [40] A. T. T. Al-Hameedi, H. H. Alkinani, S. Dunn-Norman et al., "Mud loss estimation using machine learning approach," *Journal of Petroleum Exploration and Production Technology*, vol. 9, no. 2, pp. 1339–1354, 2019.
- [41] H. H. Alkinani, A. T. T. Al-Hameedi, S. Dunn-Norman, M. M. Alkhamis, and R. A. Mutar, "Prediction of lost circulation prior to drilling for induced fractures formations using artificial neural networks," in *SPE-195197-MS was presented at SPE Oklahoma City Oil and Gas Symposium*, Oklahoma City, OKC, USA, 2019.

- [42] A. K. Abbas, N. A. Al-Haideri, and A. A. Bashikh, "Implementing artificial neural networks and support vector machines to predict lost circulation," *Egyptian Journal of Petroleum*, vol. 28, no. 4, pp. 339–347, 2019.
- [43] R. A. L. Cristofaro, G. A. Longhin, A. A. Waldmann et al., "Artificial intelligence strategy minimizes lost circulation non-productive time in Brazilian deep water pre-salt," in *OTC-28034-MS Presented at the Offshore Technology Conference OTC Brasil*, Rio de Janeiro, Brazil, 2017.
- [44] R. Jahanbakhshi and R. Keshavarzi, "Quantitative and qualitative analysis of lost circulation in natural and induced fractured formations: the integration of operational conditions and geomechanical parameters," *European Journal of Environmental and Civil Engineering*, vol. 19, no. 4, pp. 418–444, 2014.
- [45] L. Sheremetov, I. Z. Batyrshin, D. Filatov, J. Martinez, and H. Rodriguez, "Fuzzy expert system for solving lost circulation problem," *Applied Soft Computing*, vol. 8, no. 1, pp. 14–29, 2008.