The Smart Sensorial Technologies for Smart Healthcare

Lead Guest Editor: Mohamed Elhoseny
Guest Editors: Mohammad Hammoudeh and Xiaohui Yuan

The Smart Sensorial Technologies for Smart Healthcare

The Smart Sensorial Technologies for Smart Healthcare

Lead Guest Editor: Mohamed Elhoseny
Guest Editors: Mohammad Hammoudeh and
Xiaohui Yuan

Contents

Contents

Retraction

Retracted: Application Effect Analysis of Operating Room Detailed Nursing Based on Medical Big Data in Patients Undergoing Gastrointestinal Tumor Surgery

Journal of Healthcare Engineering

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

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

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

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

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

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

References

[1] Y. Meng, A. Sun, G. Ji, C. Wei, and J. Jia, "Application Effect Analysis of Operating Room Detailed Nursing Based on Medical Big Data in Patients Undergoing Gastrointestinal Tumor Surgery," *Journal of Healthcare Engineering*, vol. 2022, Article ID 8575305, 14 pages, 2022.

Retraction

Retracted: Tensiomyography and Statistical Analysis Based Muscle Change Detection in Multiple Sclerosis for Smart Healthcare

Journal of Healthcare Engineering

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

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

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

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

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

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

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

References

[1] L. Rusu, M. C. Neamtu, O. M. Neamtu et al., "Tensiomyography and Statistical Analysis Based Muscle Change Detection in Multiple Sclerosis for Smart Healthcare," *Journal of Healthcare Engineering*, vol. 2022, Article ID 5225851, 7 pages, 2022.

Hindawi

Retraction

Retracted: Correlation Meta-Analysis of the Efficacy of Inhaled Corticosteroids in Children with Asthma Based on Smart Medical Health

Journal of Healthcare Engineering

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

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

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

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

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

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

References

[1] Y. Qin, J. Wang, J. Qin et al., "Correlation Meta-Analysis of the Efficacy of Inhaled Corticosteroids in Children with Asthma Based on Smart Medical Health," *Journal of Healthcare Engineering*, vol. 2022, Article ID 6220774, 12 pages, 2022.

Retraction

Retracted: Analyzing the Patient Behavior for Improving the Medical Treatment Using Smart Healthcare and IoT-Based Deep Belief Network

Journal of Healthcare Engineering

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

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

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

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

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

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

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

References

[1] R. M. K. Mohamed, O. R. Shahin, N. O. Hamed, H. Y. Zahran, and M. H. Abdellattif, "Analyzing the Patient Behavior for Improving the Medical Treatment Using Smart Healthcare and IoT-Based Deep Belief Network," *Journal of Healthcare Engineering*, vol. 2022, Article ID 6389069, 8 pages, 2022.

Retraction

Retracted: Prenatal Monitoring of Perinatal Pregnant Women and Fetus Based on a Smart Electronic Fetal Monitoring System

Journal of Healthcare Engineering

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

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

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

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

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

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

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

References

[1] Y. Sun and S. Jiang, "Prenatal Monitoring of Perinatal Pregnant Women and Fetus Based on a Smart Electronic Fetal Monitoring System," *Journal of Healthcare Engineering*, vol. 2022, Article ID 5073636, 7 pages, 2022.

Retraction

Retracted: Advanced Cognitive Algorithm for Biomedical Data Processing: COVID-19 Pattern Recognition as a Case Study

Journal of Healthcare Engineering

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

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

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

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

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

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

References

[1] M. Elhoseny, Z. Tarek, and I. M. EL-Hasnony, "Advanced Cognitive Algorithm for Biomedical Data Processing: COVID-19 Pattern Recognition as a Case Study," *Journal of Healthcare Engineering*, vol. 2022, Article ID 1773259, 11 pages, 2022.

Hindawi

Research Article

Intelligent Energy-Aware Thermal Exchange Optimization with Deep Learning Model for IoT-Enabled Smart Healthcare

Mahmoud Ragab [ID][1,2,3] **and Sami Saeed Binyamin** [ID][4]

[1]Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[2]Department, Faculty of Science, Al-Azhar University, Nasr City 11884, Cairo, Egypt
[3]Centre for Artificial Intelligence in Precision Medicines, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[4]Computer and Information Technology Department, The Applied College, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Correspondence should be addressed to Mahmoud Ragab; mragab@kau.edu.sa

In recent years, Internet of Things (IoT) and advanced sensor technologies have gained considerable interest in linking different medical devices, patients, and healthcare professionals to improve the quality of medical services in a cost-effective manner. The evolution of the smart healthcare sector has considerably enhanced patient safety, accessibility, and operational competence while minimizing the costs incurred in healthcare services. In this background, the current study develops intelligent energy-aware thermal exchange optimization with deep learning (IEA-TEODL) model for IoT-enabled smart healthcare. The aim of the proposed IEA-TOEDL technique is to group the IoT devices into clusters and make decisions in the smart healthcare sector. The proposed IEA-TEODL technique constructs clusters using the energy-aware chaotic thermal exchange optimization-based clustering (EACTEO-C) scheme. In addition, the disease diagnosis model also intends to classify the collected healthcare data as either presence or absence of the disease. To accomplish this, the proposed IEA-TODL technique involves several subprocesses such as preprocessing, K-medoid clustering-based outlier removal, multihead attention bidirectional long short-term memory (MHA-BLSTM), and weighted salp swarm algorithm (WSSA). The utilization of outlier removal and WSSA-based hyper-parameter tuning process assist in achieving enhanced classification outcomes. In order to demonstrate the enhanced outcomes of the IEA-TEODL approach, a wide range of simulations was conducted against benchmark datasets. The simulation results inferred the enhanced outcomes of the IEA-TEODL technique over recent techniques under distinct evaluation metrics.

1. Introduction

With the advancements made in smart sensorial media, Internet of Things (IoT), and cloud techniques, smart health care has gained considerable interest in different domains such as healthcare, academia, government, and industry [1]. In recent times, Internet of Things (IoT) has brought the vision of a smart world into reality, with numerous services in the pipeline generating massive amounts of data. Cloud computing (CC) suits well as an enabling technique since it presents a flexible stack of software, computing, and storage services at a lower cost [2]. Cloud-based service has the potential to provide a high-quality seamless experience to clinicians, physicians, and other caregivers, anytime and anywhere. While research has been making advances in cloud services and IoT separately, minimum attention has been paid to emerging, affordable, and cost-effective intelligent healthcare services [3]. At present, cloud and IoT technologies have assisted in delivering smart healthcare services on a real-time basis and also have made considerable improvements.

With the integration of the IoT cloud, a great demand for intelligent and smart healthcare systems provides a rapid and seamless response. Artificial intelligence (AI) and deep

learning (DL) techniques can improve decision-making and cognitive behaviour [4]. Advanced electronic applications are presented to intelligent healthcare stakeholders along with smart sensor devices. In spite of these, it is challenging to access or find hospitals and medical professionals in intelligent healthcare environments. In general, patients with serious medical needs must be provided quick attention and faster response in order to save their lives [5]. Therefore, data recorded from patients needs to be interpreted and transferred to healthcare professionals with minimum delay while the results need to be sufficiently accurate so that it can be utilized by healthcare experts for disease prognosis. Hence, a smart healthcare system is required that could resolve the above-mentioned problems and leverage the technology and services available in the intelligent healthcare environment. Figure 1 illustrates the structure of a smart healthcare system.

Though there have been advancements in this domain, the concept of a smart healthcare system remained uncertain without cognitive function. Smart city service can never be exploited completely without the cognitive knowledge of its stakeholders [6]. Even though the conventional methods achieve rapid delivery of results, it is expected to obtain highly accurate results. But, most of the time, the results suffer from complex data [7]. In this situation, high accuracy can be accomplished by deep learning (DL) techniques and its different versions. In literature, these techniques are trained using large datasets [8]. DL method is an emerging field that has gained considerable outcomes in sequence prediction, mixed-modality data sets, and natural language processing tasks that have received heavy growth in various applications such as computer vision and speech recognition [9, 10].

The current article develops intelligent energy-aware thermal exchange optimization with deep learning (IEA-TEODL) model for IoT-enabled smart healthcare. The proposed IEA-TEODL technique derives energy-aware chaotic thermal exchange optimization-based clustering (EACTEO-C) scheme. Besides, a disease diagnosis model is also involved to classify the collected healthcare data into either presence or absence of the disease. To accomplish this, the proposed IEA-TODL technique involves several subprocesses such as preprocessing, K-medoid clustering-based outlier removal, multihead attention bidirectional long short-term memory (MHA-BLSTM), and weighted salp swarm algorithm (WSSA). In order to validate the promising performance of the IEA-TEODL technique, a wide range of simulations was performed against benchmark datasets, and the results were validated under different measures.

2. Literature Review

Mansour et al. [11] developed a disease diagnosis system for diabetes and heart disease using IoT and AI convergence methods. The presented technique employed crow search optimization approach-based cascaded LSTM (CSO-CLSTM) for disease diagnoses. To accomplish improved classification of healthcare information, CSO was employed for tuning "weights" and "bias" parameters of the presented

approach. The authors in the literature [12] developed a cloud-centric IoT-based *m*-healthcare monitoring disease diagnosis system that predicts the possible disease occurrence with the severity level. In this study, key terminology was determined to generate user-based health measurement by examining computation science concepts.

In literature [13–15], the authors presented a disease diagnosis system with DL as well as IoT. The healthcare information is preprocessed since it contains noise. The preprocessed information is then passed onto isolation forest (iForest) for outlier recognition with high precision and linear time complexity. The data undergo a classification method in which DenseNet169 and PSO methods are incorporated to diagnose the disease; the parameter is then tuned to improve the performance. Awotunde et al. [16] developed an IoT-WBN-based architecture with an ML approach. The data collected from wearable sensors such as glucose sensors, body temperature, chest, and heartbeat sensors are transferred by IoT device to the cloud dataset.

Nagarajan et al. [17] designed an IoT-based FoG-enabled cloud network framework that accumulates real-time healthcare information from patients through a number of healthcare IoT sensor networks. This information is examined by the DL technique deployed in a fog-based healthcare environment. Moreover, the presented approach was utilized in sustainable smart city solutions to estimate real-time process. Ihnaini et al. [18] proposed an intelligent healthcare system for diabetes based on deep ML and data fusion perspectives. With data fusion, the unrelated burden of computation abilities was removed, and the presented system's efficiency in terms of recommendation and prediction of this severe disease, in a precise format, was increased. At last, the ensemble ML approach was trained for predicting diabetes.

3. The Proposed Model

In this study, a novel IEA-TEODL technique has been developed to accomplish clustering and decision-making in an IoT-enabled smart healthcare environment. The proposed IEA-TEODL technique follows 2-stage processes, namely EACTEO-C-based cluster construction and optimal DL-based disease classification. The detailed working process of these two modules is elaborated in the succeeding subsections. Figure 2 displays the block diagram of the IEA-TEODL technique.

3.1. Process Involved in EACTEO-C Technique. In the primary stage, the IoT devices are placed in the healthcare environment to gather medical data from the patients. In order to achieve effectual energy utilization and data transmission to the cloud server, the EACTEO-C technique is executed to select the cluster head (CH) and construct it.

3.1.1. Overview of CTEO Algorithm. The primary aim behind the adaption of a meta-heuristic approach named thermal exchange optimization (TEO) is to cluster the nodes. The model of temperature from TEO reflects the
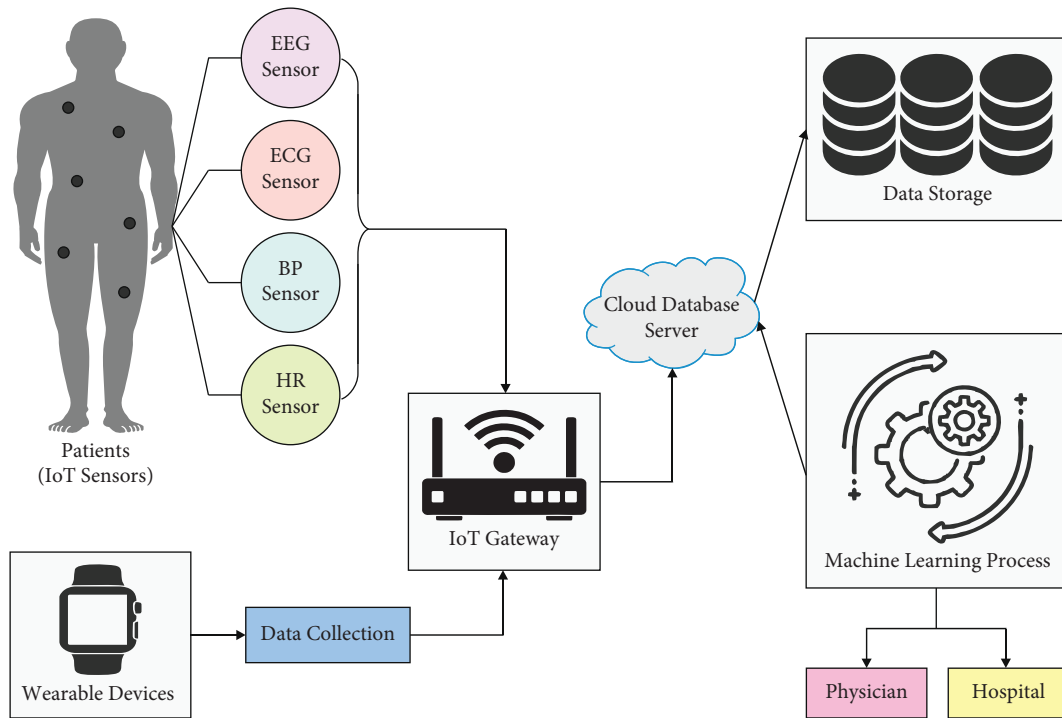
FIGURE 1: Smart healthcare systems [4].
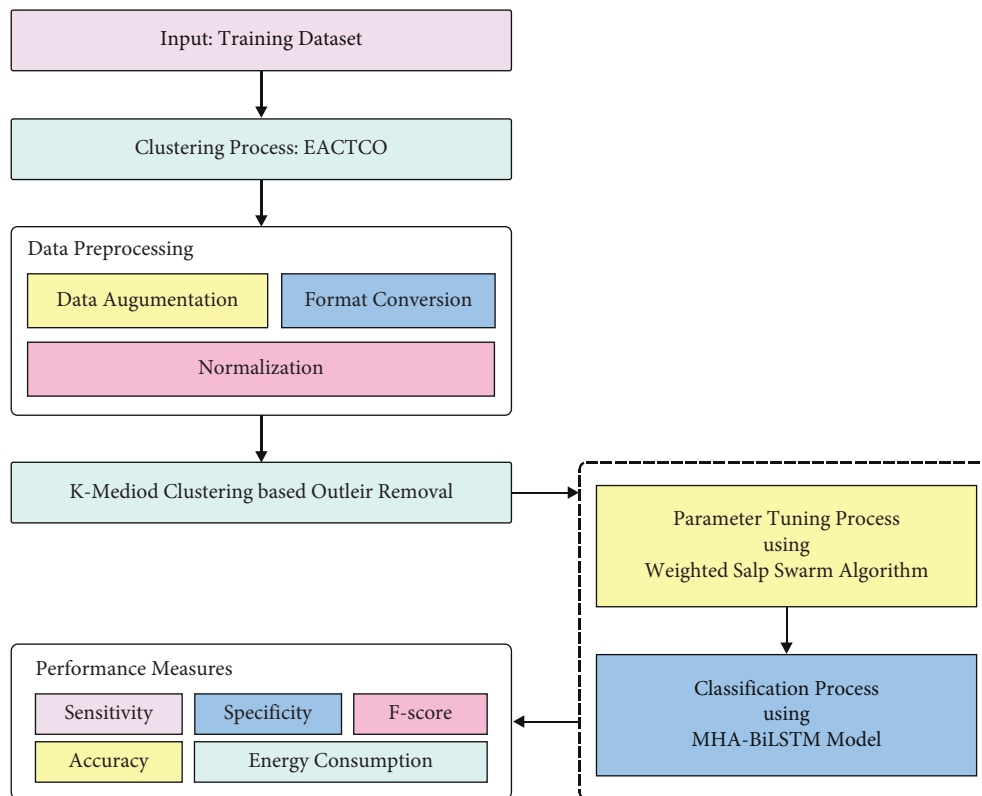
FIGURE 2: Block diagram of IEA-TEODL technique.

interface feature of nodes [19]. The cooling object mentions the place of nodes whereas the environmental temperature signifies the adjacent nodes. The object is considered as a sensor node. Therefore, important nodes are either interpreted as objects or conversely.

The primary temperature of every node is defined as follows:

$$T_i^0 = T_{\min} + \text{rndx}(T_{\max} - T_{\min}), \tag{1}$$

where T_i^0 refers to the primary solution vector of the node, i. $T_{\min}$ and $T_{\max}$ signify the limits of temperature variables. In addition, rnd stands for arbitrary vector, whereas all the components are in the range of zero and one. The main function computes the cost value of all the nodes. The memory has regarded that hierarchy holds the optimum T vector, and the main function value is connected to these vectors. It improves the technical performance with no increase in computational cost. In this way, a thermal memory (TM) is utilized to save several optimum solutions at the moment. So, during this phase, solution vectors, stored from TM, are transmitted to populations. In addition, a similar amount of accessible worse nodes is not assumed. Eventually, the node is sorted in an ascending order based on its respective main function values. The node is divided into two equivalent groups. For instance, T_1 is an environment object for $T_{n/2+1}$ cooling object and conversely.

Generally, if the β value of object is lesser, it somewhat modifies the temperatures. An analogy is simulated as this feature is projected. The value of all the nodes is calculated based on equation (2). Therefore, the β value of lesser cost node remains a minimum value, and somewhat it modifies the node place.

$$\beta = \frac{\text{cost}(\text{node})}{\text{cost}(\text{wnode})}. \tag{2}$$

The time is dependent upon the number of iterations. t denotes the time value for all the nodes and is computed as follows:

$$t = \frac{N_{\text{iter}}}{N_{\text{max}-\text{iter}}}, \tag{3}$$

where N_{iter} and $N_{\text{max}-\text{iter}}$ demonstrate the present and maximal number of iterations correspondingly. The environment temperature is replaced by equation (4). At this point, c_1 and c_2 denote control variables.

$$T_i^{\text{env}} = \left(1 - (c_1 + c_2 * (1 - t)) * \text{rnd}\right) * T_i^{p-\text{env}}. \tag{4}$$

$T_i^{p-\text{env}}$ refers to the previous temperature of the node modified to T_i^{env}.

(i) $(1 - t)$ is recognized to decrease arbitrariness when approaching the final iteration. While the procedure is nearing the end, t improves and reduces the production of arbitrariness in a linear fashion.

(ii) c_1 checks the size of arbitrary steps. Besides, c_1 contains arbitrariness if it does not utilize a descending method ($c_2 = 0$).

(iii) c_2 controls $(1 - t)$. That is, where a decrease is not needed, this could be regarded as equivalent to zero.

Where the condition of $C = 0 (c_1 = c_2 = 0)$, the preceding temperature is multiplied by I'' and c_1 and c_2 are chosen in {0 or 1}. With the preceding stages and equation (4), the upgrade temperature of all the nodes is defined based on equation.

$$T_i^{\text{new}} = T_i^{\text{env}} + (T_i^{\text{prev}} - T_i^{\text{env}})\exp(-\beta t). \tag{5}$$

P_r parameter from $(0, 1)$ defines whether the element of all the nodes is replaced. To all the nodes, P_r is related to $\text{rnd}(i) (i = 1, 2, \cdots, n)$ and is an arbitrary number that is equally distributed from zero and one. If $\text{rnd}(i) < P_r$, a dimension nodes, i is arbitrarily selected, and their values are redefined as follows:

$$T_{ij} = T_{j,\min} + \text{rndx}(T_{j,\max} - T_{j,\min}), \tag{6}$$

where $T_{i,j}$ refers to the variable j of node i. $T_{j,\min}$ and $T_{j,\max}$ imply lower as well as upper limits of the variable j correspondingly. Only one size is altered to preserve the infrastructure of nodes. This method presents many benefits to nodes for moving throughout the searching region and attaining the optimum diversity.

In this work, the TEO algorithm can be improved with the design of the CTEO algorithm using chaotic concepts [20]. A chaos map employs chaotic variables with changeable nature before arbitrary variables. This order is initiated from nonlinear and dynamic systems whereas nonconvergent orders are from nonperiodic and bounded systems. It can offer easy searching together with a superior convergence rate than arbitrary search. This process uses the technique for providing the optimum exploration from solution spaces due to their dynamic performance of turbulence sequence. The current analysis utilizes a sinusoidal chaotic map function to improve both convergence speed and premature convergence of the TEO technique so as to consider a trade-off between exploitation as well as exploration techniques. This is performed to provide a well-defined outcome from the solution space which does not stuck at the local optimum points. In order to modify the TEO approach with the help of a chaos map, the chaos value is replaced with arbitrary numbers using the important formula as follows:

$$r_{i+1} = P.r_i^2 \sin(\pi.r_i), \tag{7}$$

where P defines the control parameter, and r_i and r_{i+1} imply the chaotic arbitrary numbers generated from preceding and the existing iterations correspondingly. At this point, $r_0 = 0.7$ and $P = 2.3$.

3.1.2. Application of EACTEO-C Technique for CH Selection.
The primary goal of the EACTCO-C technique is to minimize the distance among the carefully chosen CH nodes. The main objective is to minimize the delay during the transmission of information from one node to another. In contrast, for the network energy should be higher, it should

consume a small number of energies at the time of data communication. The objective function of the adapted CH is given in equation (7), where the value of η must depend upon $0 < \eta < 1$. Now, v_m and v_n show the operations as given as follows. The constraints on distance, delay, and energy are stated as σ_1, σ_2, and σ_3. The condition of this constraint is represented by $\sigma_1 + \sigma_2 + \sigma_3 = 1$. $X^x - B_s$ represents the distance between normal and sink nodes.

$$H_n = \eta v_n + (1 - \eta)v_{\mathbf{m}}, \tag{8}$$

$$v_m = \sigma_1 * v_{i^{\text{dis}}} + \sigma_2 * v_{i^{\text{ene}}} + \sigma_3 * v_{i^{\text{del}}},$$

$$v_n = \frac{1}{b} \sum_{x=1}^{b} \|X^x - B_s\|, \tag{9}$$

where $v_{(m)}^{\text{dis}}$ represents the packet transmission from the normal node to CH and from CH to BS. $v_{i^{\text{dis}}}$ must depend upon $[0, 1]$. The value of $v_{i^{\text{dis}}}$ remains high when the normal node is more along with distance among CH [21].

$$v_{i^{\text{dis}}} = \frac{v_{(\mathbf{m})}^{\text{dis}}}{v_{(n)}^{\text{dis}}},$$

$$v_{(m)}^{\text{dis}} = \sum_{x=1}^{N_x} \left[\|C_x - B_s\| + \sum_{y=1}^{N_y} \|C_x - X_x\| \right], \tag{10}$$

$$v_{(n)}^{\text{dis}} = \sum_{x=1}^{N_\chi} \sum_{y=1}^{N_y} \|X_x - X_y\|.$$

X_x denotes the normal node in x^{th} cluster, C_x represents the CH of xth cluster, the distance between the BS and CH is shown as $C_x - B_s$, $C_x - X_x$ represents the distance between normal node and CH, and $X_x - X_y$ shows the distance among two normal nodes, N_x and N_y indicate the node amount that does not assume x^{th} and y^{th} cluster. The value of $v_{i^{\text{ene}}}$ becomes higher than one, and the whole CH cumulative $v_{(m)}^{\text{ene}}$ and $v_{(n)}^{\text{ene}}$ is considered as less energy value with high number of CH s.

$$v_i^{\text{ene}} = \frac{v_{(m)}^{\text{ene}}}{v_{(n)}^{\text{ene}}}. \tag{11}$$

Delta fitness function is directly proportionate to each node that resides in the cluster. Thus, a delay gets reduced, when the CH owns a lesser number of nodes. The denominator N_N shows the overall number of nodes in WSN, and the numerator indicates the high amount of CH. Furthermore, the value of v_i^{del} must be in $d[0, 1]$.

$$v_i^{\text{del}} = \frac{\max(\|C_x - X_x\|)_{x=1}^{N_c}}{N_N}. \tag{12}$$

3.2. Disease Diagnosis Module.
In this work, the disease diagnosis model encompasses a series of subprocesses, namely preprocessing outlier removal, MHA-BLSTM-based classification, and WSSA-based hyperparameter optimization.

3.2.1. Data Preprocessing. At the initial stage, preprocessing takes place in different ways, namely data normalization, data transformation, and data augmentation. In this work, min-max normalization approach is used to normalize the input medical data. Besides, data are also transformed into a useful format, and data augmentation is applied using SMOTE technique to increase the size of the dataset.

3.2.2. K-Medoid Clustering. Next to data preprocessing, the outlier removal process is carried out using the K-medoid clustering approach. The K-means approach that utilizes and determines the means of data point in the calculation is mainly sensitive to the outlier. To resolve this, a new approach was developed in which the medoids are utilized rather than the average value from the cluster. Medoids are centre points from the cluster, and the approach is named as k-medoids clustering. Even though k-medoids computationally increase their demands, the k-medoids cluster is not mainly sensitive to the existence of outlier points and is appropriate to discrete and continuous fields of information [22]. Generally, the input provided has the value of k that denotes the amount of clusters determined to data. For every k cluster, a k-reference point is chosen. The variance between k-medoids and k-means algorithms is that the former k-medoids considers the point as a reference object for the cluster whereas k-means considers the average value from the former k-medoid cluster as the reference point.

3.2.3. Data Classification Using MHA-BLSTM Model. During the data classification process, the MHA-BLSTM model can be employed for the classification process. RNN is a well-known technique to train the series data, namely image processing, video capture, and word prediction that could remember the series element using a memory cell. The main problem of handling RNN is that once it is utilized for training with long step size, it cannot remember the data for a longer period since the backpropagated gradient either shrinks or grows at every time step. This makes the training weight vanish or explode. LSTM memory overcomes this problem while a standard LSTM unit consists of input, output, and forget gates that control the data into and out of the memory cell. The structure of a single LSTM cell includes the logistic sigmoid function whereas i, f, 0, and c represent the input gate, forget gate, output gate, and cell state, correspondingly. The input gate determines the ratio of input and has an impact on the value of the cell state [23]. The framework could resolve the exploding and vanishing gradient problems.

Figure 3 demonstrates the framework of Bi-LSTM. Bi-LSTM has both forward and backward LSTM layers. The forward layer captures the historical data of order while the backward layer captures the future data of the sequences. The combined layers are linked to a similar resultant layer. Our network utilizes Bi-LSTM with a multihead (MH) process. MH permits the model for combined data to appear in various representations of subspaces at distinct places. The attention process plays a vital role in the DL
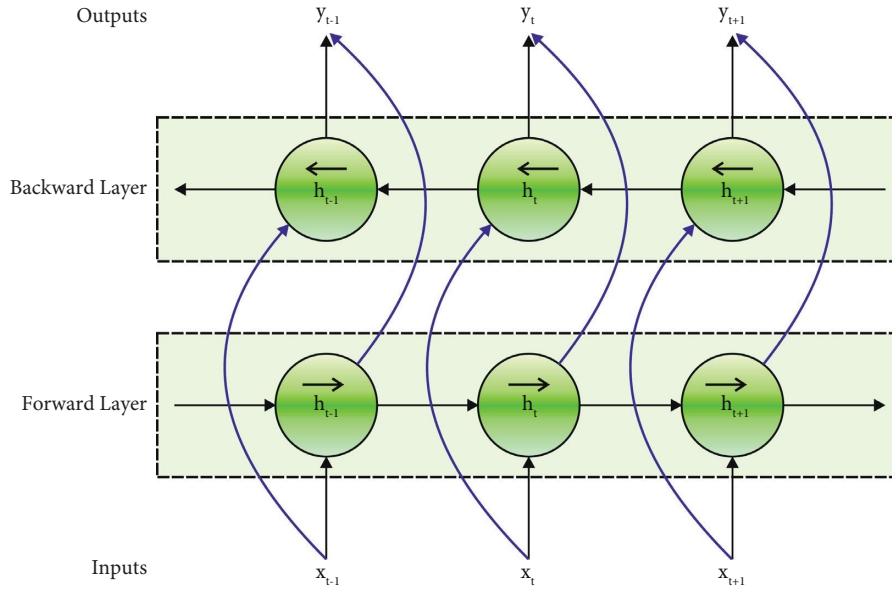
Figure 3: Bi-LSTM structure [23].

network to capture the explicit and latent context. MH attention process is presented since it utilizes several individual attention functions to capture distinct contexts. The attention function gets input as an order of query $Q = \{Q_1, \ldots, Q_N\}$ and group of key-value pairs $\{K, V\} = \{(K_1, V_1), \ldots, (K_R, V_R)\}$. MH attention method primary transforms Q, K, and V to C subspaces, with distinct and learnable linear projection.

At this point. Q^c, K^c, and V^c signify the cth head of query, key, and value correspondingly. $\{W_c^Q, W_c^K, W_c^V\} \in \mathbb{R}^{d \times d_k}$ implies the parameter matrices, and d and d_k stand for models and their subspace dimensions. Moreover, C attention functions are executed concurrently to obtain the resultant state, $O^1, \ldots, O^C$.

$$O^c = A^c V^c,$$
$$A^c = \text{softmax}\left(\frac{Q^c K^{c^T}}{\sqrt{d_k}}\right). \tag{13}$$

A^c implies the attention distribution, formed by c^{th} attention head. These resultant states are concatenated to produce the last state.

3.2.4. Parameter Tuning Using WSSA Technique. In order to fine-tune the parameters involved in the DL model, the WSSA technique is used which in turn improves the classifier results. The SSA approach is stimulated from navigation behaviour of salps in search of food in the ocean [24]. It is classified as leader and follower. In the searching method of optimization technique, it is important to balance the exploration and exploitation capabilities to accomplish better efficiency. The idea of inertia weight factor is initially presented to quicken the convergence speed. Researchers find that when inertia weight is lesser, the particle has stronger exploitation capability. However, it easily falls into local optima. In contrast, when inertia weight is larger, the particle still has a stronger exploration ability; however, the searching efficacy becomes low. Furthermore, the researcher presented the inertia weight factor to enhance the searching method. Here, the weight factor reduces linearly to balance between exploration and exploitation ability; thus, the particle has a stronger global searching capability in the earlier stage and searches for the precise outcome in the later stage. In the current study, to enhance the outcomes from traditional SSA, a weight factor is included to update the position. It changes dynamically with the number of iterations [25]. The weighted factor decreases linearly with the number of iterations from maximum to minimum values to accomplish optimal outcomes.

$$w(t) = w\text{maX} - \left(\frac{(w\max - w\min)^\star t}{L}\right), \tag{14}$$

whereas $w\max$ and $w\min$ denote the maximal and minimal values of the weighted factors, t represents the present iteration, and L indicates the maximal iteration. The position is upgraded in WSSA for leader and follower and is modelled as follows:

$$X^1 = \begin{cases} w^\star F + c_1\left((\text{UB} - \text{LB}) * c_2 + LB\right) c_3 \geq 0, \\ w^\star F - c_1\left((\text{UB} - \text{LB}) \star c_2 + LB\right) c_3 \geq 0, \end{cases} \tag{15}$$
$$X^j = 0.5 * w * \left(X^j + X^{j-1}\right),$$

whereas the variable has a similar meaning as in SSA.

WSSA approach derives a fitness function to accomplish better classification accuracy. It describes a positive integer to characterize the improved accuracy of the candidate solution. Here, the minimization of the classification error rate is taken into account as the fitness function. The optimum solution has the least error rate whereas the worst solution achieves an increased error rate.

$$\text{fitness}(x_i) = \text{classifier error rate}(x_i)$$

$$= \frac{\text{number of misclassified instances}}{\text{Total number of instances}} * 100.$$

$$(16)$$

4. Experimental Validation

In this section, the proposed IEA-TEODL model is experimentally validated for its performance using a heart disease dataset [26]. It comprises of 270 samples with 13 attributes such as age, sex, chest pain value, resting blood sugar, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, old peak, slope of peak exercise, number of major vessels, and thal. Besides, the dataset includes two class labels, namely the presence of CKD and the absence of CKD.

4.1. Results Analysis. Table 1 and Figure 4 provide the overall results of the analysis of the IEA-TEODL model on the heart disease dataset under three runs. The results demonstrate that the proposed IEA-TEODL model accomplished an effectual classification outcome under all runs. For instance, with run-1, the IEA-TEODL model achieved a $sens_y$ of 98.76%, $spec_y$ of 93.09%, $accu_y$ of 91.27%, and an F_{score} of 95.61%. Along with that, with run-2, the proposed IEA-TEODL approach accomplished a $sens_y$ of 98.21%, $spec_y$ of 92.56%, $accu_y$ of 94.19%, and an F_{score} of 94.16%. In line with these, with run-3, IEA-TEODL methodology offered a $sens_y$ of 99.15%, $spec_y$ of 96.32%, $accu_y$ of 95.92%, and an F_{score} of 99.33%.

Figure 5 depicts the ROC curve generated by the IEA-TEODL approach under three runs. The figure exposes that the proposed IAOA-DLFD technique reached an enhanced outcome with maximum output under different runs. For the sample, with run-1, the proposed IEA-TEODL methodology obtained a high ROC of 97.0602. Likewise, with run-2, the IEA-TEODL algorithm obtained an enhanced outcome (ROC) of 97.4922. Eventually, with run-3, the proposed IEA-TEODL system achieved an increased ROC of 98.4221.

Figure 6 provides the accuracy and loss graph analysis results accomplished by the IEA-TEODL approach under three runs. The outcomes show that the accuracy value increased while the loss value decreased with an increase in epoch count. It can be also understood that the training loss is low, and validation accuracy is high under three runs.

4.2. Discussion. A brief $sens_y$ analysis was conducted on the IEA-TEODL model against existing ones, and the results are shown in Table 2 and Figure 7. The results report that the proposed IEA-TEODL model achieved better outcomes in terms of $sens_y$ under distinct instances. For instance, with 2000 instances, IEA-TEODL model reached an increased $sens_y$ of 96.58%, but NN approach, NB methodology, SVM system, and ANN models obtained reduced $sens_y$ values

TABLE 1: Analytical results of IEA-TEODL technique under three runs.

No. of runs	Sensitivity	Specificity	Accuracy	F-score
Run-1	98.76	93.09	91.27	95.61
Run-2	98.21	92.56	94.19	94.16
Run-3	99.15	96.32	95.92	99.33
Average	98.71	93.99	93.79	96.37
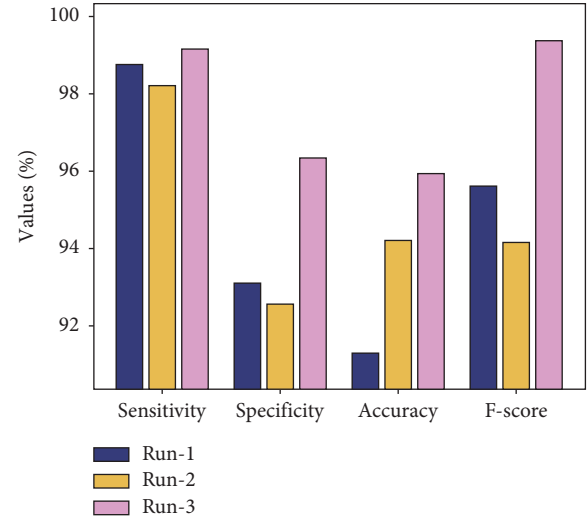
FIGURE 4: Analytical results of IEA-TEODL technique under three runs.

such as 93.55%, 87.97%, 83.16%, and 95.33% correspondingly. In addition, with 10000 instances, the proposed IEA-TEODL model reached an increased sen of 99.15%, while NN approach, NB methodology, SVM system, and ANN models obtained reduced $sens_y$ values such as 93.47%, 88.26%, 84.21%, and 98.70%, respectively.

A comparative $spec_y$ analysis was conducted on the IEA-TEODL model against existing ones, and the results are shown in Table 3 and Figure 8. The results report that the proposed IEA-TEODL approach achieved better outcomes in terms of $spec_y$ under various instances. For instance, with 2000 instances, IEA-TEODL approach reached an increased $spec_y$ of 95.40%, whereas NN approach, NB methodology, SVM system, and ANN models obtained the least $spec_y$ values such as 84.86%, 83.71%, 80.93%, and 94.36% respectively. Furthermore, with 10000 instances, the proposed IEA-TEODL technique reached an increased $spec_y$ of 96.32%, whereas NN approach, NB methodology, SVM system, and ANN methodologies obtained less $spec_y$ values such as 90.26%, 86.91%, 84.13%, and 91.90% correspondingly.

A detailed acc_y analysis was conducted on the IEA-TEODL algorithm against existing methods, and the results are shown in Table 4 and Figure 9. The results report that the proposed IEA-TEODL technique achieved better outcomes with respect to acc_y under distinct instances. For instance, with 2000 instances, the proposed IEA-TEODL model attained an increased acc_y of 94.28%, but NN approach, NB methodology, SVM system, and ANN systems obtained less
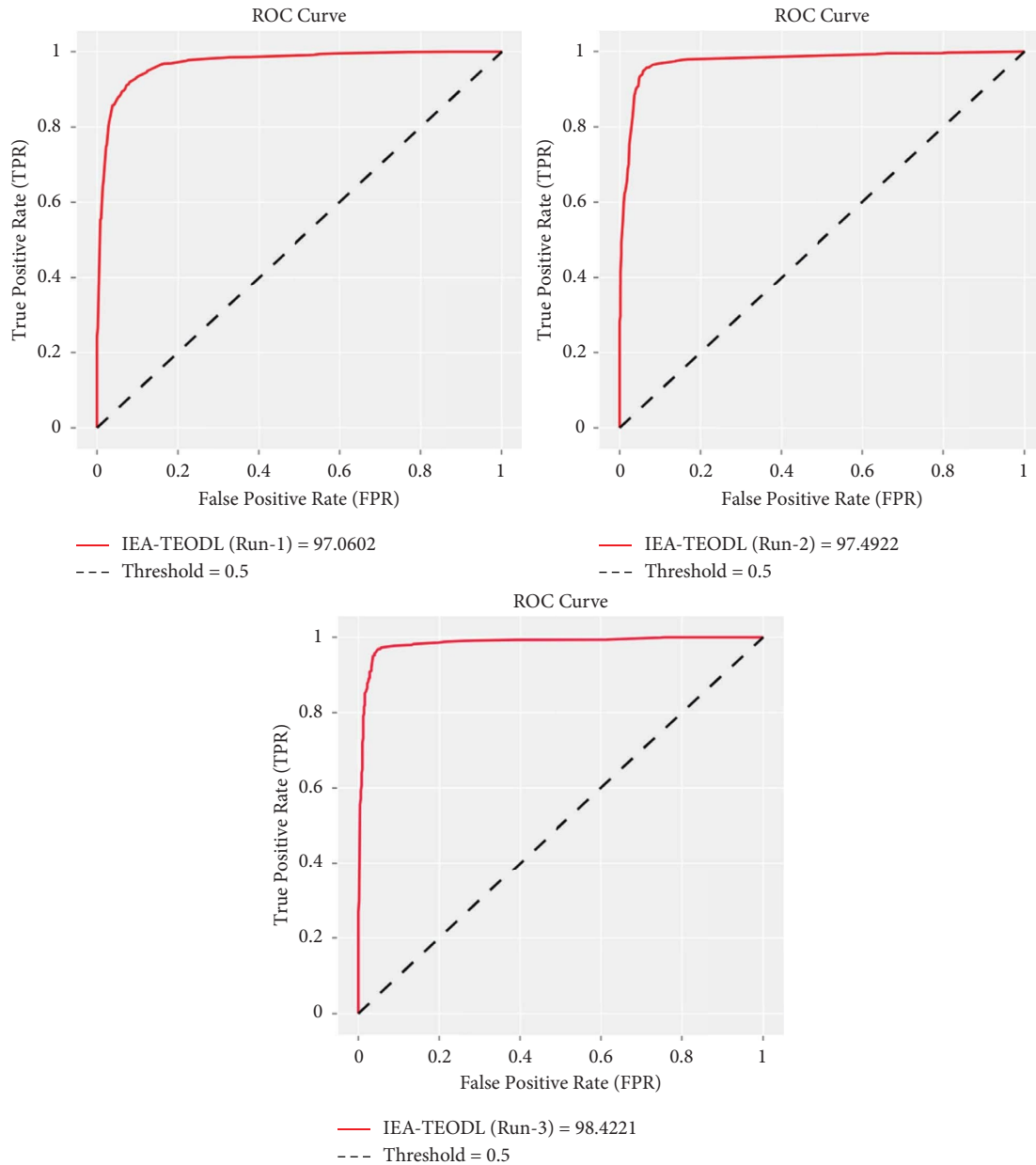
FIGURE 5: ROC analysis results of IEA-TEODL technique under three runs.

acc_y values such as 88.73%, 77.43%, 73.17%, and 92.54% correspondingly.

Additionally, with 10000 instances, the proposed IEA-TEODL approach reached the maximum acc_y of 95.92%, whereas NN approach, NB methodology, SVM system, and ANN models obtained low acc_y values, namely 89.61%, 82.02%, 81.98%, and 93.88% correspondingly.

A brief F_{score} analysis was conducted between the IEA-TEODL method and the existing models, and the results are shown in Table 5 and Figure 10. The results infer that the proposed IEA-TEODL approach achieved better outcomes in terms of F_{score} under distinct instances. For instance, with 2000 instances, the presented IEA-TEODL model reached the maximum F_{score} of 98.32%, while NN approach, NB methodology, SVM system, and ANN algorithms obtained

low F_{score} values such as 92.33%, 84.63%, 81.59%, and 97.67% correspondingly. Finally, with 10000 instances, the proposed IEA-TEODL algorithm obtained an increased F_{score} of 99.33%, whereas NN approach, NB methodology, SVM system, and ANN models reached less F_{score} values such as 97.71%, 84.25%, 82.32%, and 95.84% correspondingly.

At last, a brief TEC examination was conducted between IEA-TEODL model and recent methods, and the results are shown in Table 6 and Figure 11 [27]. The experimental values highlight that the proposed IEA-TEODL model produced effective TEC values under distinct IoT sensor counts. For instance, with 100 IoT sensors, the IEA-TEODL model gained a low TEC of 41.30%, whereas EE-PSO, ABC, GWO, and ACO algorithms obtained high TEC values such
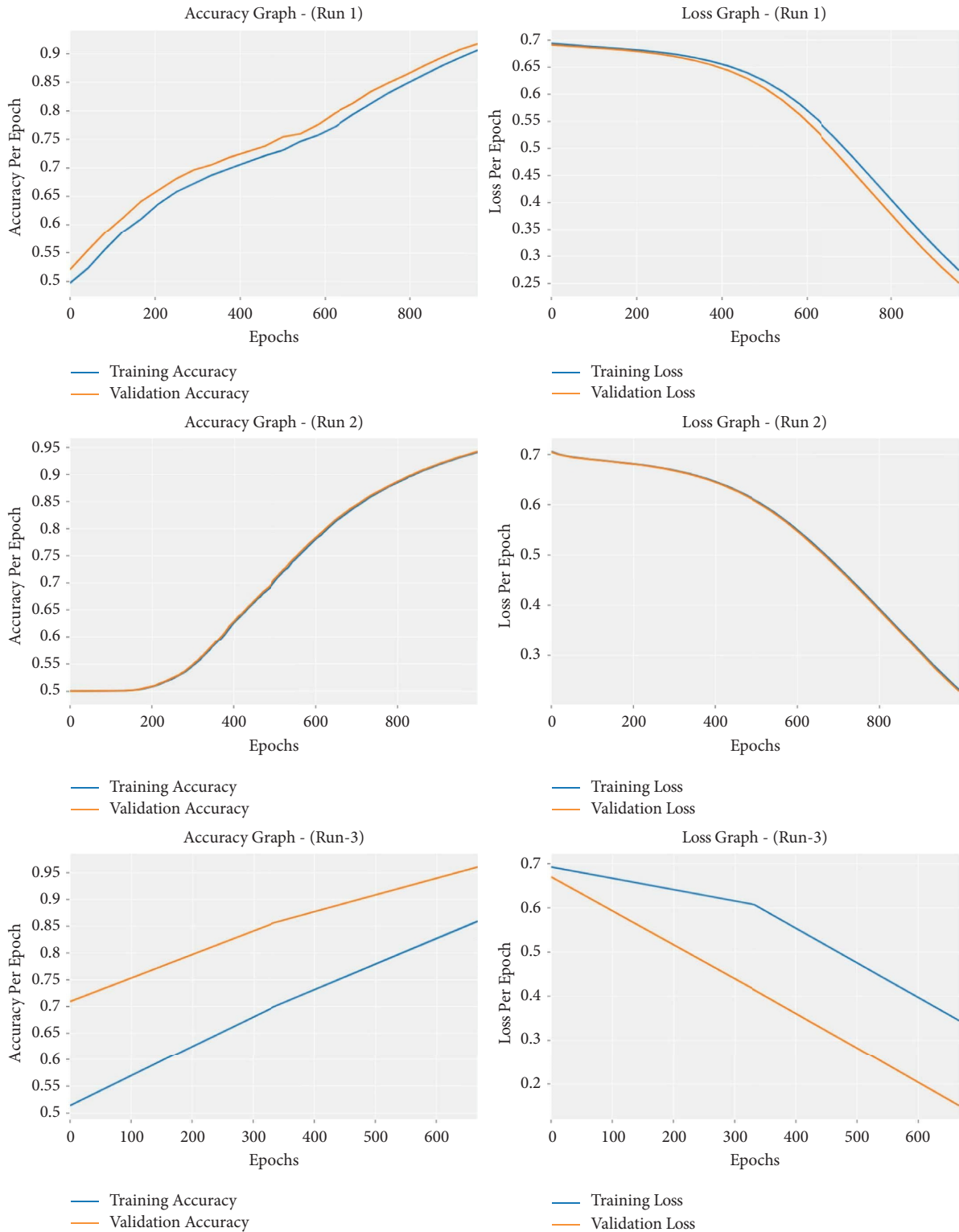
FIGURE 6: Accuracy and loss analysis results of IEA-TEODL technique under three runs.

as 45.04%, 57.14%, 60.65%, and 66.16%, respectively. At the same time, with 300 IoT sensors, the proposed IEA-TEODL method gained a low TEC of 57.71%, whereas EE-PSO, ABC, GWO, and ACO systems obtained high TEC values such as 59.73%, 67.24%, 73.44%, and 77.15% correspondingly. In line with this, with 500 IoT sensors, the proposed IEA-TEODL model gained a low TEC of 65.74%, whereas EE-PSO, ABC, GWO, and ACO approaches attained high TEC values namely 69.28%, 78.51%, 82.11%, and 84.08% correspondingly.

TABLE 2: Sensitivity analysis results of IEA-TEODL technique against existing approaches.

Instances	Nearest neighbour	Naive Bayes	SVM model	ANN model	IEA-TEODL
2000	93.55	87.97	83.16	95.33	96.58
4000	88.03	85.03	81.95	94.33	95.41
6000	92.92	86.87	83.62	95.53	97.41
8000	92.02	88.98	81.40	96.96	98.71
10000	93.47	88.26	84.21	98.70	99.15
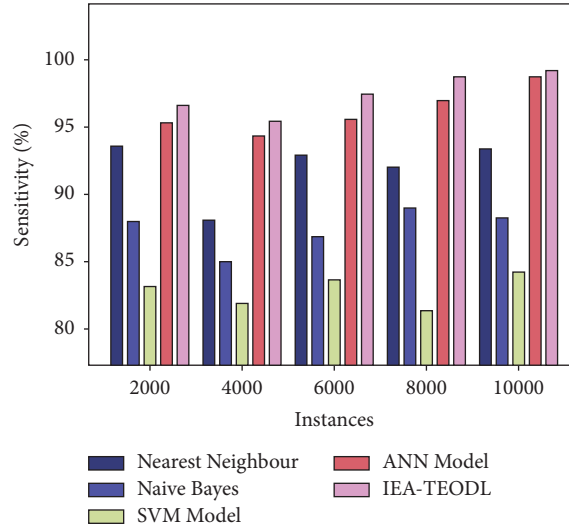
FIGURE 7: $sens_y$ analysis of the IEA-TEODL technique with recent approaches.
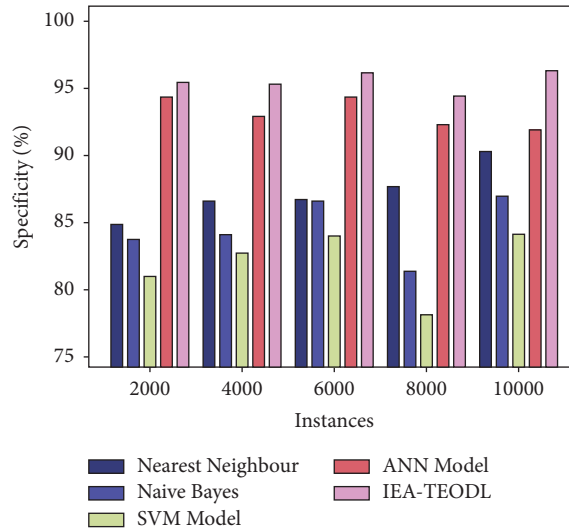
FIGURE 8: $Spec_y$ analysis results of the IEA-TEODL technique against recent approaches.

TABLE 3: Specificity analysis results of IEA-TEODL technique against existing approaches.

Instances	Nearest neighbour	Naive Bayes	SVM model	ANN model	IEA-TEODL
2000	84.86	83.71	80.93	94.36	95.40
4000	86.58	84.08	82.75	92.92	95.29
6000	86.72	86.59	84.01	94.27	96.14
8000	87.65	81.34	78.14	92.26	94.40
10000	90.26	86.91	84.13	91.90	96.32

TABLE 4: Accuracy analysis results of the IEA-TEODL technique against existing approaches.

Instances	Nearest neighbour	Naive Bayes	SVM model	ANN model	IEA-TEODL
2000	88.73	77.43	73.17	92.54	94.28
4000	90.97	77.80	76.69	94.33	95.86
6000	86.99	76.89	75.86	92.59	93.70
8000	86.43	80.42	77.86	94.82	95.15
10000	89.61	82.02	81.98	93.88	95.92
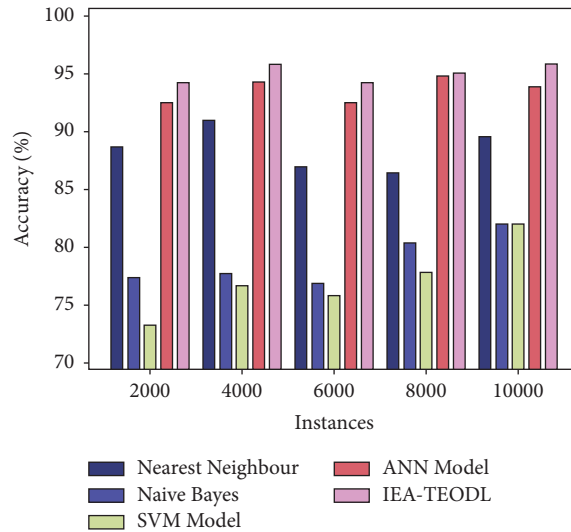
FIGURE 9: Acc_y analysis results of the IEA-TEODL technique against recent approaches.
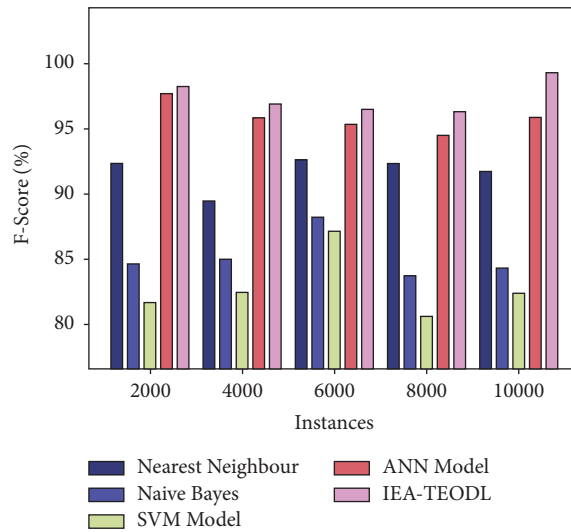
FIGURE 10: F_{score} analysis results of the IEA-TEODL technique against recent approaches.

TABLE 5: F-score analysis results of the IEA-TEODL technique against existing approaches.

Instances	Nearest neighbour	Naive Bayes	SVM model	ANN model	IEA-TEODL
2000	92.33	84.63	81.59	97.67	98.32
4000	89.47	84.91	82.45	95.87	96.87
6000	92.57	88.19	87.16	95.33	96.52
8000	92.28	83.73	80.61	94.49	96.29
10000	91.71	84.25	82.32	95.84	99.33

TABLE 6: Results of the analysis of total energy consumption (%) between the existing and the proposed methods.

IoT sensors	IEA-TEODL	EE-PSO model	ABC model	GWO model	ACO model
100	41.30	45.04	57.14	60.65	66.16
200	49.85	52.35	61.01	68.66	74.26
300	57.71	59.73	67.24	73.44	77.15
400	60.99	63.72	72.49	76.00	81.06
500	65.74	69.28	78.51	82.11	84.08
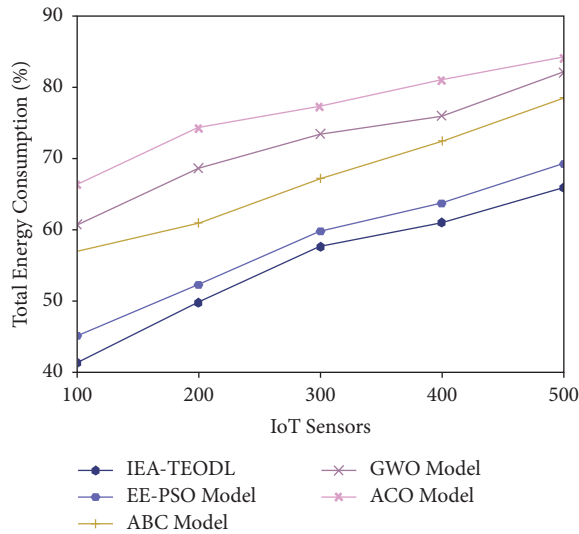
FIGURE 11: TEC analysis of IEA-TEODL technique with recent approaches.

After examining the above-mentioned tables and figures, it is apparent that the proposed IEA-TEODL technique outperformed other methods. The enhanced performance of the proposed model is due to the integration of EACTEO-C-based cluster construction and optimal DL-based disease classification.

5. Conclusion

In this study, a novel IEA-TEODL technique has been developed to accomplish clustering and decision-making in an IoT-enabled smart healthcare environment. The proposed IEA-TEODL technique follows a two-stage process, namely EACTEO-C-based cluster construction and optimal DL-based disease classification. Besides, the disease diagnosis model encompasses a series of subprocesses, namely preprocessing outlier removal, MHA-BLSTM-based classification, and WSSA-based hyperparameter optimization. In order to validate the promising performance of the proposed IEA-TEODL technique, a wide range of simulations was conducted against benchmark datasets. The simulation results established the enhanced outcomes of the IEA-TEODL technique over other recent techniques under distinct evaluation metrics. Thus, the IEA-TEDOL technique can be utilized as an effectual tool to accomplish energy efficiency and data classification in an IoT environment. In the future, lightweight cryptography and authentication mechanisms can be included to assure security in the smart healthcare environment.

Data Availability

Data sharing is not applicable to this article as no datasets were generated during the current study.

Consent

Not applicable.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] T. M. Ghazal, M. K. Hasan, M. T. Alshurideh et al., "IoT for smart cities: machine learning approaches in smart healthcare—a review," *Future Internet*, vol. 13, no. 8, p. 218, 2021.

[2] M. Islam, A. Rahaman, and M. R. Islam, "Development of smart healthcare monitoring system in IoT environment," *SN computer science*, vol. 1, no. 3, pp. 185–211, 2020.

[3] F. Alshehri and G. Muhammad, "A comprehensive survey of the Internet of Things (IoT) and AI-based smart healthcare," *IEEE Access*, vol. 9, pp. 3660–3678, 2021.

[4] M. V. V. Kantipudi, C. J. Moses, R. Aluvalu, and S. Kumar, "Remote patient monitoring using IoT, cloud computing and AI," in *Hybrid Artificial Intelligence and IoT in Healthcare*, pp. 51–74, Springer, 2021.

[5] R. Patan, G. S. Pradeep Ghantasala, R. Sekaran, D. Gupta, and M. Ramachandran, "Smart healthcare and quality of service in IoT using grey filter convolutional based cyber physical system," *Sustainable Cities and Society*, vol. 59, Article ID 102141, 2020.

[6] T. Hussain, K. Muhammad, S. Khan, A. Ullah, M. Y. Lee, and S. W. Baik, "Intelligent baby behavior monitoring using embedded vision in IoT for smart healthcare centers," *Journal of Artificial Intelligence and Systems*, vol. 1, no. 1, pp. 110–124, 2019.

[7] G. Liu, Y. Fan, J. Zhang, P. Wen, Z. Lyu, and X. Yuan, "Deep flight track clustering based on spatial–temporal distance and denoising auto-encoding," *Expert Systems with Applications*, vol. 198, Article ID 116733, 2022.

[8] M. Poongodi, A. Sharma, M. Hamdi, M. Maode, and N. Chilamkurti, "Smart healthcare in smart cities: wireless

patient monitoring system using IoT," *The Journal of Supercomputing*, vol. 77, no. 11, pp. 12230–12255, 2021.

[9] S. Sengan, O. I. Khalaf, S. Priyadarsini, D. K. Sharma, K. Amarendra, and A. A. Hamad, "Smart healthcare security device on medical IoT using raspberry pi," *International Journal of Reliable and Quality E-Healthcare*, vol. 11, no. 3, pp. 1–11, 2021.

[10] S. Mian Qaisar and A. Subasi, "Effective epileptic seizure detection based on the event-driven processing and machine learning for mobile healthcare," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 7, pp. 3619–3631, 2020.

[11] R. F. Mansour, A. E. Amraoui, I. Nouaouri, V. G. Díaz, D. Gupta, and S. Kumar, "Artificial intelligence and Internet of Things enabled disease diagnosis model for smart healthcare systems," *IEEE Access*, vol. 9, pp. 45137–45146, 2021.

[12] P. Verma and S. K. Sood, "Cloud-centric IoT based disease diagnosis healthcare framework," *Journal of Parallel and Distributed Computing*, vol. 116, pp. 27–38, 2018.

[13] E. A. Refaee and S. Shamsudheen, "A computing system that integrates deep learning and the internet of things for effective disease diagnosis in smart health care systems," *The Journal of Supercomputing*, vol. 78, no. 7, pp. 9285–9306, 2022.

[14] G. Madhu, B. Lalith Bharadwaj, R. Boddeda et al., "Intelligent disease diagnosis model for energy aware cluster based IoT healthcare systems," *Computers, Materials & Continua*, vol. 71, no. 1, pp. 1189–1203, 2022.

[15] M. M. Althobaiti, K. Pradeep Mohan Kumar, D. Gupta, S. Kumar, and R. F. Mansour, "An intelligent cognitive computing based intrusion detection for industrial cyber-physical systems," *Measurement*, vol. 186, Article ID 110145, 2021.

[16] J. B. Awotunde, S. O. Folorunso, A. K. Bhoi, P. O. Adebayo, and M. F. Ijaz, "Disease diagnosis system for IoT-based wearable body sensors with machine learning algorithm," in *Hybrid Artificial Intelligence and IoT in Healthcare*, pp. 201–222, Springer, 2021.

[17] S. M. Nagarajan, G. G. Deverajan, P. Chatterjee, W. Alnumay, and U. Ghosh, "Effective task scheduling algorithm with deep learning for internet of health things (ioht) in sustainable smart cities," *Sustainable Cities and Society*, vol. 71, Article ID 102945, 2021.

[18] B. Ihnaini, M. A. Khan, T. A. Khan et al., "A smart healthcare recommendation system for multidisciplinary diabetes patients with data fusion based on deep ensemble learning," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 4243700, 11 pages, 2021.

[19] A. Kaveh and A. Dadras, "A novel meta-heuristic optimization algorithm: thermal exchange optimization," *Advances in Engineering Software*, vol. 110, pp. 69–84, 2017.

[20] L. Hu, Y. Zhang, and N. Yousefi, "Nonlinear modeling of the polymer membrane fuel cells using deep belief networks and modified water strider algorithm," *Energy Reports*, vol. 7, pp. 2460–2469, 2021.

[21] D. L. Reddy, C. Puttamadappa, and H. N. Suresh, "Merged glowworm swarm with ant colony optimization for energy efficient clustering and routing in Wireless Sensor Network," *Pervasive and Mobile Computing*, vol. 71, Article ID 101338, 2021.

[22] D. Syed, H. Abu-Rub, A. Ghrayeb et al., "Deep learning-based short-term load forecasting approach in smart grid with clustering and consumption pattern recognition," *IEEE Access*, vol. 9, pp. 54992–55008, 2021.

[23] S. R. B. Shah, G. S. Chadha, A. Schwung, and S. X. Ding, "A sequence-to-sequence approach for remaining useful lifetime estimation using attention-augmented bidirectional LSTM," *Intelligent Systems with Applications*, vol. 10-11, Article ID 200049, 2021.

[24] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp Swarm Algorithm: a bio-inspired optimizer for engineering design problems," *Advances in Engineering Software*, vol. 114, pp. 163–191, 2017.

[25] N. Chouhan, U. R. Bhatt, and R. Upadhyay, "Weighted salp swarm and salp swarm algorithms in FiWi access network: a new paradigm for ONU placement," *Optical Fiber Technology*, vol. 63, Article ID 102505, 2021.

[26] archive, "Statlog (Heart)," 2022, https://archive.ics.uci.edu/dataset/145/statlog+heart.

[27] R. Bharathi, T. Abirami, S. Dhanasekaran et al., "Energy efficient clustering with disease diagnosis model for IoT based sustainable healthcare systems," *Sustainable Computing: Informatics and Systems*, vol. 28, Article ID 100453, 2020.

Retraction

Retracted: Tensiomyography and Statistical Analysis Based Muscle Change Detection in Multiple Sclerosis for Smart Healthcare

Journal of Healthcare Engineering

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

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

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

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

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

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

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

References

[1] L. Rusu, M. C. Neamtu, O. M. Neamtu et al., "Tensiomyography and Statistical Analysis Based Muscle Change Detection in Multiple Sclerosis for Smart Healthcare," *Journal of Healthcare Engineering*, vol. 2022, Article ID 5225851, 7 pages, 2022.

Research Article

Tensiomyography and Statistical Analysis Based Muscle Change Detection in Multiple Sclerosis for Smart Healthcare

Ligia Rusu [iD],[1] **Marius Cristian Neamtu** [iD],[2] **Oana Maria Neamtu** [iD],[1] **Mihai Robert Rusu** [iD],[1] **Mihnea Ion Marin** [iD],[3] **Daniel Danciulescu** [iD],[1] **and Jude Hemanth** [iD][4]

[1]*Sport Medicine and Physical Therapy Department, University of Craiova, Craiova 200585, Romania*
[2]*Physiology Department, University of Medicine and Pharmacy Craiova, Craiova 200349, Romania*
[3]*Applied Mechanic Department, University of Craiova, Craiova 200585, Romania*
[4]*Department of ECE, Karunya Institute of Technology and Sciences, Coimbatore, India*

Correspondence should be addressed to Jude Hemanth; judehemanth@karunya.edu

The impact of demyelinization on muscle fiber changes and the type of changes in multiple sclerosis (MS) is very hard to estimate. One of the major problems of MS patients is muscle fatigue and decrease of muscle force in the range of 16–57%. The objective of this research work is to estimate various aspects of muscle changes at tibial muscle (mTA) level using a noninvasive method named as tensiomyography (TMG). TMG provides information about muscle functions in MS. This study includes 40 MS patients among which 18 are males (45%) and 22 are females (55%). They are divided in two subgroups: subgroup A and subgroup B. Subgroup A includes 20 MS patients without clinical decelable gait disorders and subgroup B includes 20 MS patients with clinical decelable gait disorders. Also, we have a control group that includes 20 healthy people with the same average age. Average age is 38.15 ± 11.19 y for MS patients and 39.34 ± 10.57 for healthy people. Evaluation measures include ADL score and EDSS scale. The ADL score is 0 for patients from subgroup A and 1 for patients from subgroup B. The EDSS score is 1 for subgroup A and 2.5 for subgroup B. This study confirms the importance of TMG based evaluation of muscle changes in MS patients. This smart healthcare system is also used for prediction of the muscle changes and muscle imbalance. Contraction time (Tc) recordings are used to detect the muscle fatigue which is a specific symptom of MS. The value of Tc for subgroup A is 45.8 ms and subgroup B is 61.37 ms for right side. Analysis of these two parameters such as Dm and Tc could define the muscle behaviour and help provide early information about the possibility of developing gait disorders. This smart TMG system analyses the muscle tone in the best possible way to predict the onset of any diseases which is an integral part of the smart healthcare system.

1. Introduction

Multiple sclerosis (MS) is an autoimmune disease which affects myelin that surrounds nerve cell axons. The result of the demyelinization process is specific for MS which leads to the development of the local inflammation and subsequently the installation of scars nerve cell axons. This results in the decrease of nervous conductibility. This is the main reason for the sensor-motor difficulties in MS patients. However, the demyelinization process is very complex and nonlinear. It is very difficult to estimate the intensity of muscle fiber changes and the type of changes. Several researchers have

analyzed the decrease of muscle force with the isokinetic system which leads to muscle atrophy [1, 2]. Specifically, lower limb muscles are much affected than the upper limb muscles [3].

One of the major problems for MS patients is muscle fatigue and decrease of muscle force that could be between 16 and 57% [4]. The main reason seems to be the demyelinization process at the upper central neuron [5]. Accordingly, muscle activation in MS is reduced because of the deficiency in motor units [4]. Other observations are the decrease in the production of peripheric nervous mediators, reduction in the dimensions of muscle fibers, change of

muscle types fibers distribution, and neuromuscular transmission. Muscle changes-based research in MS also includes muscle biopsy at vastus lateral muscle. MS patients with EDSS score of 4.75 show the presence of type II-a myosin heavy chain (MHC) isoform. This increases in according with the disability level and is associated with decrease of muscle fiber type I.

Thus, fibers from the MS group displayed a subtle shift in fast MHC isoform coexpression and a modest reduction in cross-bridge number, density, or average force. All these occurs with no change in maximal cross-bridge cycling rate or susceptibility to intracellular metabolites. These changes explain the part of the muscle weakness and fatigue experienced by individuals with MS [6]. MS patients seem to have a lower percentage of type I fiber and a higher percentage of type II fiber. Changes in muscle fibers distribution are connected with spasticity and muscle weakness which are seen in most of the MS patients. In spite of the availability of many literature on spasticity, the necessity for more information still persists for the assessment of muscle weakness in MS. Muscle weakness is one of the main features that disturbs the muscle force which impacts the daily activities of human beings.

2. Related Works

In this context, there are a lot of studies that discuss about correlation between muscle weakness in MS and the MMT scale (manual muscle testing) based estimated muscle force values. McDonald and Compston, in their book McAlpine's Multiple Sclerosis [7], show that from a group of 301 patients, 52% suffer from muscle weakness. Hoang et al. demonstrates the presence of muscle weakness among 72% patients from a total of 142 patients with MS [8].

At the same time, muscle weakness in MS is associated with decrease of fitness, functional status [9, 10], and general fatigue [9]. This leads to decrease in the level of daily activities and also the quality of life. Thus, it is evident that the muscle fiber dimension reduction is the major element in the development of muscle weakness.

Literature survey describes numerous pathogenic mechanisms which are the reasons for the evolution of unpredictable nature of lesion evolution in MS patients. The unpredictability is mainly due to the presence of older lesions and new lesions at the same time. This indicates the dynamic evolution of these diseases and the nonlinear changes in the progression of this disease. Silent lesions are a specific feature and pathognomonic for MS [11]. In the context of evolution of MS, clinical evaluation is the first process and it is followed by analysis of cerebrospinal fluid. Thus, it is necessary and possible to have the evolution pattern of immune response. Evoked potential and MRI also assist in the complete diagnosis [12].

Evoked potential includes visual auditive and somatosensorial potentials for assessing the magnitude and response time after stimulation. These potentials provide a clear picture about nervous central system status functionality. The advancement in smart healthcare systems aid in the identification, development, and implementation of a new model for data acquisition for early diagnosis, monitoring, and treatment planning. The aim of this paper is to present various aspects of muscle changes at tibial muscle (mTA) level using a noninvasive method named tensiomyography (TMG) which gives us numerous information about muscle function in MS. The rationale of this study is to present some aspects regarding the possible muscle changes produced in MS with a technology like TMG. This will assist in detecting the onset of any gait disorders developed during the MS. Analysis of the muscle response after electrical stimulation with TMG provides information about the muscle fibers' response in terms of muscle fatigue. Early detection of gait abnormalities is extremely important for proper treatment planning.

The rest of the paper is organized as follows: Section 3 deals with the materials and methods, Section 4 demonstrates the statistical analysis carried out on the real-time data, Section 5 illustrates the experimental results and discussion, and Section 6 provides the conclusions with the key findings of this real-time research work.

3. Materials and Methods

The study includes 40 MS patients among which are 18 males (45%) and 22 females (55%). The control group includes 20 healthy people with the same average age for the comparative analysis of TMG parameters. However, the analysis is always complex due to the polymorphism of lesions in MS and the clinical and functional evolution. The criteria for selection of MS patients are certainty of MS diagnosis in accordance with clinical and paraclinic criteria and clear evidence for not associated with any other pathologies. Diagnosis criteria [13–15] that are accounted for in this work are as follows: certainty of MS, minimal two episodes of clinical manifestation of MS, clinical symptoms for two separately different lesions or clinical symptoms for one clinical lesion, and another subclinical lesion that is demonstrated by neurophysiologic evaluation or MRI. The exclusion criteria of this research work are lack of all clinical information about the patient and the patients that do not agree to participate in the study.

The small number of the patients is mainly because of the huge diversity in the evolution of the MS in different moments, difficulty in generating the evolution of symptomatology, and difficulty in ambulation. The patient's selection is based on gender, age, and evolution stage because of the requirement for the variety of abnormalities. However, MS is more frequent in female patients. Additionally, the following aspects are considered:

(i) All patients are from urban zone

(ii) Average age is 38.15 ± 11.19 y for MS patients and 39.34 ± 10.57 for healthy people

(iii) Subgroup A consists of 20 MS patients without clinical decelable gait disorders

(iv) Subgroup B consists of 20 MS patients with clinical decelable gait disorders

Analysis of gait disorders from the clinical point of view is made using ADL scale [16]. The interpretation is given as follows: 0 = normal (8 points); 1 = mild disability (6 points); 2 = moderate disability (4 points); 3 = severe disability (less than 4 points). Patients from subgroup A have ADL score 0, and patients from subgroup B have ADL score 1. For functional evaluation, we use EDSS scale which has a score from 0 to 10. The score is 1 for subgroup A and 2.5 for subgroup B. For subgroup A, Tc value is 60.3 ms for left side and 45.8 ms for right side. For subgroup B, Tc value is 61 ms for left side and 61.375 ms for right side. For subgroup A, Dm value is 3.2 mm for left side and 3.9 mm for right side. For subgroup B, Dm value is 3.83 mm for left side and 3.69 mm for right side. We observe a significant difference for both subgroups in terms of Tc and Dm.

Tensiomyography is a new method for the assessment of the muscle functional status which allows to evaluate the contractile proprieties of muscle fibers after electrical stimulation. The assessment includes tibial anterior muscle (mTA) and analysis of two TMG parameters: muscle displacement (Dm) and time of contraction (Tc) during electrical stimulation. These parameters are correlated with muscle fatigue and muscle composition of type I and type II muscle fibers. The statistical analysis is focused on data distribution, descriptive analysis, and analysis of significant variance ($p < 0.05$) between MS subgroups and control group. Results show normal distribution of the data and significant differences between TMG parameters that includes decrease of Dm and increase of Tc for subgroups A and B in comparison to the control group.

The method consists of application of a progressive electrical stimulation (single stimulation but progressive increase of intensity), using surface electrodes, from 10 mA until the muscle response becomes maximal. An illustration on muscle contraction is given in Figure 1. The stimulation consists of successive impulses and the frequency of stimulation is 5–25 Hz, 40–50 V. Time of stimulation is 10 s. The muscle response is collected by a sensor of TMG (G40, RLS Inc.), which is placed on the maximal muscle belly point. The exact point is detected by maximal isometric contraction in rest position before stimulation. The surface electrodes are placed towards the sensor. The stimulus is applied on the proximal electrode. The electrical stimulation generates isometric muscle contraction. The sensor has a bow of size 0.17 N/mm and is placed perpendicular to the muscle surface.

The electrical stimulation produces a transversal displacement of muscle fibers which is taken over by the sensor. The displacement is proportional to the muscle force during isometric contraction. By this way, displacement and other TMG parameters can help assess the muscle fatigue. Measure of muscle response, database, and analysis of the recording data are made using a dedicated TMG software.

The TMG parameters are as follows:

(i) Delay time (Td): the time elapsed from the moment of stimulation to the level of reaching 10% muscle contraction (ms);
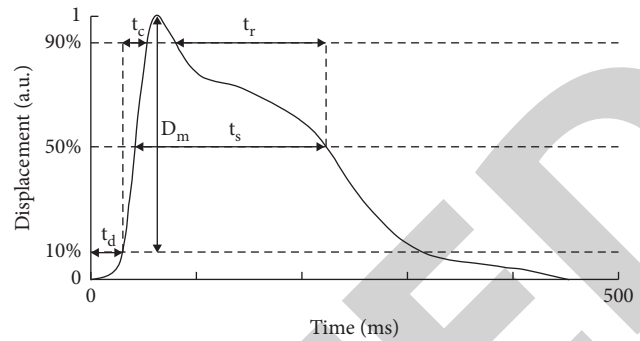
Figure 1: Graphic of muscle displacement (and other TMG parameters) during electrical stimulation.

(ii) Contraction time (Tc): the time between the moment when the muscle contraction is 10% and muscle contraction is 90% from maximal level (ms).

Contraction time depends on muscle composition (percent of type I fiber and type II fiber). The Tc increases if the percent of type II fibers increase and type I fibers decrease and this is in correlation with the onset of muscle fatigue.

(i) Sustain time (Ts): the time between the moment when the contraction is 50% and the moment when the relaxation reaches 50% (ms);

(ii) Relax time (Tr): the time elapsed from the moment when the relaxation is 50% and the moment when the relaxation is 90% (ms);

(iii) Amplitude of muscle displacement Dm (mm) is correlated with Tc and depends on muscle tissue elasticity. Dm increases during muscle explosive force and decreases if the muscle tone increases.

In this work, the analysis is performed only with two parameters Dm and Tc. These two parameters allow estimating the muscle composition and behaviour of muscle fatigue. The muscle fatigue is correlated with Tc and Dm. The assessment of tibial muscle mTA (Figure 2) is given much emphasis because this muscle is involved in foot dorsal flexion during the first moment of the gait.

4. Statistical Analysis

The aim of the statistical analysis is to identify the significant differences of Dm and Tc between the control group and subgroups of MS patients. The statistical analysis is also used for evaluating the significant differences from left and right side, for each subgroup A and B. This is to assess the functional asymmetry and muscle imbalance. Data analysis-based software packages are used for the statistical analysis.

A database is initially created with the experimental data from the significant aspects for this research which are extracted. The recording values of the parameters are analyzed to visualize the variables, and the statistical analysis is performed to observe the significant differences between data series for each group. Descriptive data (means SD) are

Figure 2: Position of the electrodes and sensor on the mTA.

Table 1: Jarque–Bera test and p value* for Dm and Tc.

JB test	m TA right side		m TA left side	
	Tc (ms)	Dm (mm)	Tc (ms)	Dm (mm)
JB test control group	1.548	1.465	1.270	1.270
P value control group	0.461	0.481	0.530	0.530
JB test group A	3.012	2.970	1.117	0.988
P value group A	0.222	0.227	0.572	0.610
JB test group B	0.055	0.041	6.016	1.208
P value group B	0.973	0.979	0.054	0.547

*Statistical significance was set at a level of $p < 0.05$.

reported for the entire patient cohort. Normal distribution is tested using the JB (Jarque–Bera) test and visual analysis of Gauss function for Dm and Tc.

The Jarque–Bera test is the easiest way to test the assumption that the values in a dataset are normally distributed. This test uses the following hypotheses: H_0: the data are normally distributed; H_1: the data are not normally distributed.

We apply the t-test for equal and unequal variances depending on the results of the Levene test, for Dm and Tc.

Levene's test is an inferential statistic used to assess the equality of variances for a variable calculated for two groups. Levene's test is used before a comparison of means. The t test is a statistical test that is used to compare the means of two groups. It is used in hypothesis testing to determine two groups that are different from one another from the TMG parameters. Statistical significance is set at a level of $p < 0.05$.

5. Experimental Results and Discussion

The normalization of the measurements for all three groups (control, subgroup A, and subgroup B) is checked using the Jarque–Bera test (JB test). Table 1 shows the analysis of Gauss function for Dm and Tc.

Statistical significance is set at a level of $p < 0.05$. It is evident from the table that all data follow normal distribution. Gauss graphics are presented in Figures 3–5 and the normal distribution is followed in them too.

Table 2 shows the average values and SD of Tc and Dm for mTA. This falls under the category of descriptive statistical analysis.

We observe in Table 2 that Dm are mostly with small values which are close for both subgroups A and B (subgroup A 3.9 mm and subgroup B 3.69 mm) for the right side. The values are lesser than those for the control group where a value of 5 mm is observed for Dm of the right side. The differences between MS subgroups and control group are also significant for the left side. The Dm values are almost similar for subgroups A and B (subgroup A 3.2 mm and subgroup B 3.83 mm). The Dm value of the control group is 4.5 mm. Tc parameters are observed with higher values for

subgroups A and B, but with a significant difference between the subgroups as shown in Table 2. The value of Tc for subgroup A is 45.8 ms and subgroup B is 61.37 ms for right side. The values are lesser than the values for the control group whose value is 42.78 ms.

However, higher values are observed for the left side for both subgroups A and B. The values are also similar with subgroup A recording a value of 60.3 ms and subgroup B recording a value of 61 ms. The values are higher than the values for the control group whose value is 41.85 ms. An analysis on the variation of Dm and Tc for subgroups A and B is carried out to check if there are any significant differences between subgroups and control group. The t-test based on series dispersion, equal, or unequal variances is used as the statistical method for this analysis. Tables 3 and 4 show the results of the Levene test which provides information about the type of dispersion.

Based on the above results, the p value is estimated to quantify the significance of the variance. These details are supplied in Tables 5–7.

A significant difference between two subgroups A and B for both parameters is observed, which is constant for both parameters. Also, significant differences between subgroups and control group for both parameters are also observed from Tables 5 and 6. An analysis is also performed for the estimation of functional symmetry (right/left) for each subgroup and the results are presented in Table 7. A significant difference is observed for Tc parameters and only for subgroup A ($p = 0.025$).

Analysis of the results in this research reveals that the average Dm values are smaller for subgroups A and B in comparison to the average values for the control group. This means that the patients from subgroups A and B have an increase of muscle tone. Tc is higher for MS patients, and it is correlated with muscle fibers type. This provides information about the decrease in the percent of muscle type I fibers and increase in the risk of muscle fatigue. Increase in the percent of muscle type II fibers is correlated with small values of Dm which indirectly indicates the increase in muscle tone. Motor performances of lower limb depend on the muscle proprieties. TMG assists in assessing the muscle changes from the morphofunctional point of view in MS patients. Muscle tone and muscle force are components that depend on the muscle composition and muscle atrophy. Muscle atrophy is often visible in MS patients, which is evidence for the onset of muscle imbalance.
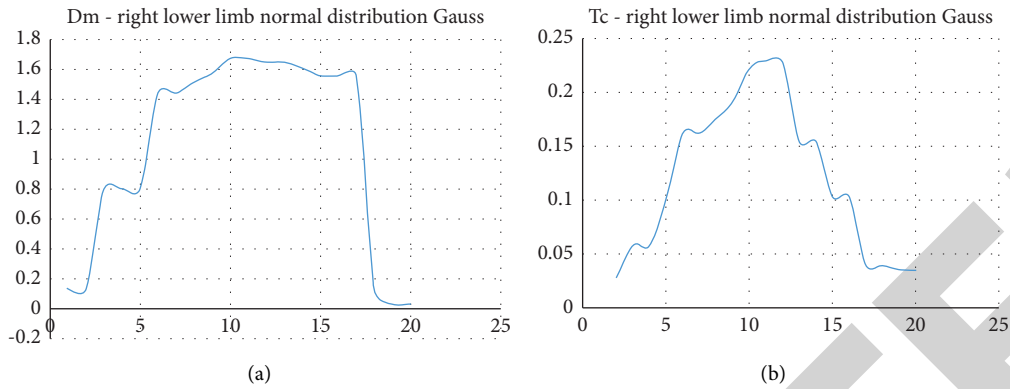
Figure 3: Control group Gauss function: (a) Dm; (b) Tc.
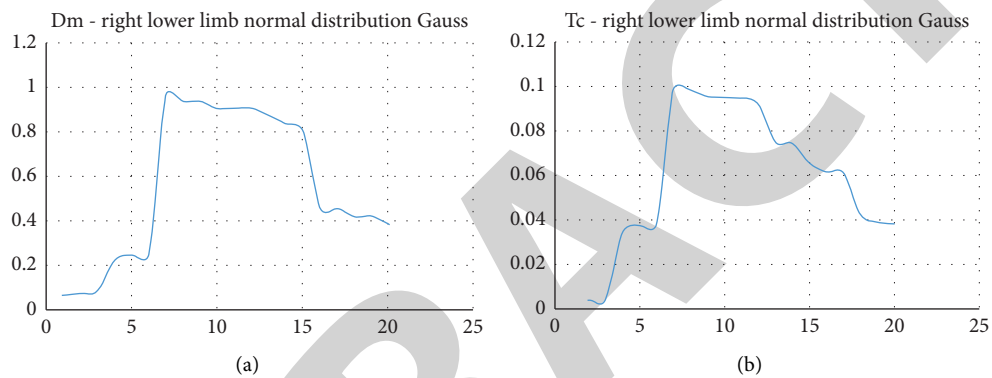
Figure 4: Subgroup A Gauss function: (a) Dm; (b) Tc.
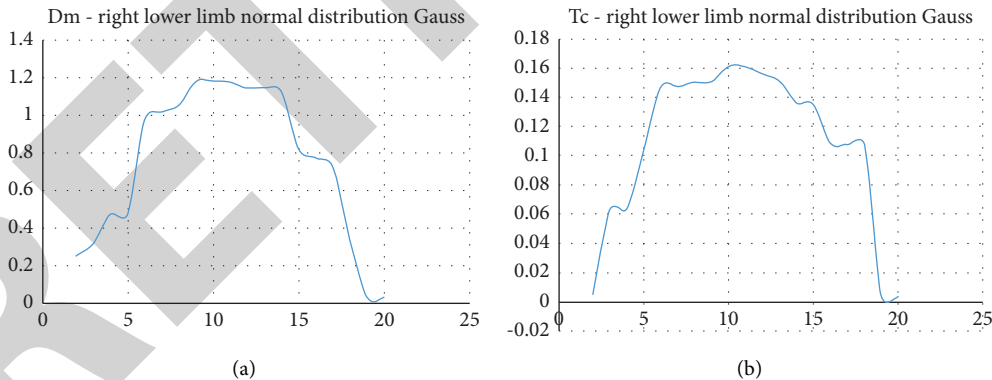
Figure 5: Subgroup B Gauss function: (a) Dm; (b) Tc.

The small values of Dm in MS and the values of Dm for subgroup A explain the nondecelable clinical gait disorders of subgroup A. This aspect is related with the ankle kinetic during the gait and it is an important element for initiation of the gait. In subgroup B, the average value for Dm is very less than the values of the control group. This is the main reason for the gait disorders to be clinically complex between subgroups A and B. Tc parameter values are higher and this suggests an increase of muscle fibers type II. This could be considered like an adaptive process during the pathology of

MS. This aspect was also observed by Kent-Braun et al. [17] in their study. At the same time, muscle biopsy of mTA revealed the decrease in the percent of type I fibers and increase in type II fibers in MS patients. This new muscle configuration is called as "fiber effect" and is based on fibril IIx (transition fibers). It also depends on the morpho-functional muscle changes and requirements.

Most of the research works have focused only on muscle changes during spasticity in stroke. In our research, more emphasis is given for the assessment using TMG for MS

Table 2: Average values and standard deviation (SD) for Dm and Tc, for m TA.

Group		Dm (mm)		Tc (ms)	
		Right	Left	Right	Left
Control group	Average	5	4.5	42.78	41.85
	SD	2.1	2.2	10.69	11.70
MS without gait disorders (subgroup A)	Average	3.9	3.2	45.8	60.3
	SD	2.3	2.3	22.57	22
MS with gait disorders (subgroup B)	Average	3.69	3.83	61.375	61
	SD	1.45	1.91	14.82	18

Table 3: Result of the Levene test for Tc.

TA muscle	Subgroups A-B	Control group – subgroup A	Control group – subgroup B
Right	t-UV	t-UV	t-EV
Left	t-UV	t-UV	t-EV

t-UV, t-test unequal variances; t-EV, t-test equal variances.

Table 4: Result of the Levene test for Dm.

TA muscle	Subgroups A-B	Control group – subgroup A	Control group – subgroup B
Right	t-EV	t-UV	t-EV
Left	t-EV	t-EV	t-EV

t-UV, t-test unequal variances; t-EV, t-test equal variances.

Table 5: P value* (test t-Student) for Tc, compare subgroups A and B, compare subgroups and control group.

TA muscle	Subgroups A-B	Control group – subgroup A	Control group – subgroup B
Right	0.008	0.034	0.00038
Left	0.044	0.01	0.0001

*Statistical significance was set at a level of $p < 0.05$.

Table 6: P value* (test t-Student) for Dm, compare subgroups A and B, compare subgroups and control group.

TA muscle	Subgroups A-B	Control group – subgroup A	Control group – subgroup B
Right	0.046	0.044	0.013
Left	0.047	0.017	0.0182

*Statistical significance was set at a level of $p < 0.05$.

Table 7: P values* (test t-Student) for Dm and Tc, functional symmetry right/left.

TA muscle	Right/left – subgroup A	Right/left – subgroup B
Tc	0.025	0.485
Dm	0.158	0.210

*Statistical significance was set at a level of $p < 0.05$.

patients. This is the main contribution of the work. The proposed model is in contrast to the study of Krizaj et al. [18] which analyses the TMG parameters (Tr, Ts, Dm) in stroke for monitoring the effect of botulin toxin BTX-A administration. This study confirms the importance of TMG in the assessment process of muscle changes in MS, monitoring the evolution and the prediction of the muscle changes and muscle imbalance. Based on the study of Šimunic et al. [19], this work can be extended in future to propose the rehabilitation process in athletes in terms of muscle fatigue.

6. Conclusions

A smart healthcare system is heavily dependent on the smart diagnostics methodologies. In this research work, the smart sensor based TMG system is used to detect the muscle strength, which can be used to predict the gait disorders and other MS abnormalities. The smart assessment is based on Tc and Dm. Contraction time Tc is high for mTA, and it is correlated with increase of type II fibers and muscle fatigue which are specific symptoms of MS. TMG is a noninvasive way for assessment of the muscle proprieties without involving the tendon property or joint movement during the electrical stimulation. Analysis of these two parameters such as Dm and Tc can define the muscle behaviour and help to have early information about the possibility of developing the gait disorders. Hence, the therapeutic intervention can be started early to limit the progression of muscle damage

Hindawi

Retraction

Retracted: Prenatal Monitoring of Perinatal Pregnant Women and Fetus Based on a Smart Electronic Fetal Monitoring System

Journal of Healthcare Engineering

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

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

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

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

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

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

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

References

[1] Y. Sun and S. Jiang, "Prenatal Monitoring of Perinatal Pregnant Women and Fetus Based on a Smart Electronic Fetal Monitoring System," *Journal of Healthcare Engineering*, vol. 2022, Article ID 5073636, 7 pages, 2022.

Research Article

Prenatal Monitoring of Perinatal Pregnant Women and Fetus Based on a Smart Electronic Fetal Monitoring System

Yu Sun and Su'e Jiang [ID]

Huai'an Maternal and Child Health Care Hospital, Huai'an 221006, China

Correspondence should be addressed to Su'e Jiang; 161847355@masu.edu.cn

The aim of the study is to study the prenatal monitoring of perinatal pregnant women based on a smart electronic fetal monitoring system. Through the comparative analysis of 230 pregnant women in maternal and child health care hospital who received fetal heart monitoring during the perinatal period and those who did not receive fetal heart monitoring during the perinatal period, cases of fetal distress, neonatal asphyxia, and cesarean section were observed in both groups. Results show that the incidences of fetal complications and cesarean sections in the experimental group were 16.36% and 36.82%, which was significantly higher than 4.50% and 17.50% in the control group($p < 0.05$); the neonatal mild and severe asphyxia rates in the experimental group were 3.18% and 1.36%, which were significantly lower than 9.50% and 6.50% in the control group ($p < 0.05$). The experimental results show that the correct application of fetal heart rate monitoring in the perinatal period can aid in early detection and dealing with fetal distress and reduce the occurrence of various complications such as neonatal asphyxia. It is worthy to be popularized and applied in clinics.

1. Introduction

Fetal distress caused by intrauterine hypoxia is an important factor leading to fetal death. According to statistics, at least 30% of newborns born every year have cerebral palsy and 10% of patients with severe mental retardation are caused by fetal intrauterine hypoxia. The fetal heart is an organ that supplies oxygen and transports nutrients. It is controlled by the central nervous system of the brain and regulated by body fluids such as blood flow hormones (Figure 1). If there are problems, it can lead to fetal distress, growth retardation, even premature delivery, and dystocia [1]. However, these clinical features can be reflected from the changes of the fetal heart rate and uterine contractility (CTG) curve. CTG is a kind of monitor that describes the fetal heart rate and uterine contractile pressure of pregnant women and is also called a heart delivery force recorder. Different fetal heart contractions (CTG) have a complete set of different interpretation methods. The traditional interpretation method is to analyze the fetal heart contraction curve with the naked eye and experience. In recent years, the

electronic fetal heart rate monitoring system has developed rapidly. With the introduction of computer-aided analysis of characteristics of the fetal heart rate cardiotocograph (CTG) curve and clinical parameters, fetal heart rate changes due to anemia or hypoxia can be detected early. Therefore, a smart electronic fetal monitoring system is of great significance for eugenics, early detection of fetal abnormalities, and prevention of fetal damage [2, 3]. At present, it is the most commonly used, sensitive, and effective monitoring method of fetal intrauterine conditions. Especially for the third trimester of pregnancy, effective fetal electronic monitoring can timely find the existence of fetal distress, and then through active and effective intervention, the incidence and mortality of various complications such as neonatal asphyxia can be significantly reduced.

2. Related Works

Gomez, O., pointed out through research that the fetus should be monitored after 32 weeks of pregnancy. On the one hand, it can ensure high security. On the other hand, it
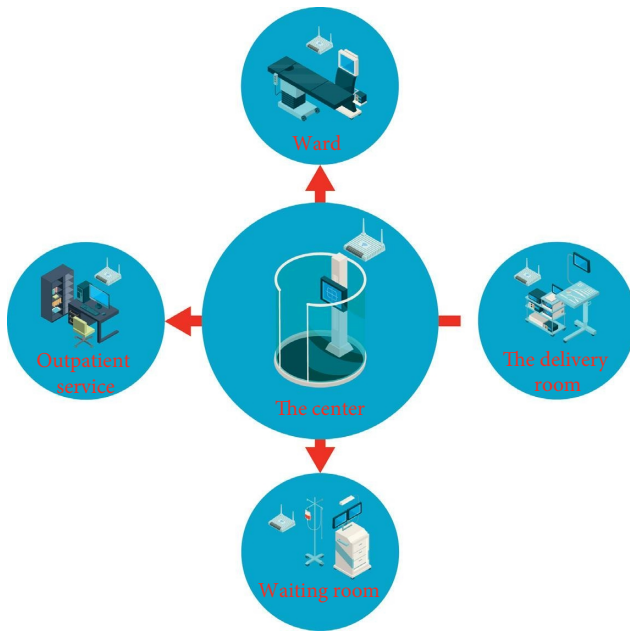
Figure 1: Fetal detection system.

can minimize and avoid a series of adverse results caused by iatrogenic over prediction. [4]. Han, J., and others suggested that antenatal fetal monitoring should be initiated at 32–34 weeks of gestation. If pregnant women had high-risk factors such as diabetes, the start-up time should be advanced to 26–28 weeks. We chose the time to start fetal heart rate monitoring for pregnant women without complications at 36 weeks of pregnancy [5]. Khan, S. M., and others conducted a prenatal fetal heart monitoring on 780 pregnant women and screened 756 cases of the reactive type and 41 cases of the nonreactive type. The rate of reactive cesarean section was 25.1% (196/756), and the rate of nonreactive cesarean section was 61.0% (25/41) [6]. Amjad O. performed NST once a week for 1400 pregnant women after 36 weeks of pregnancy. The value of NST in predicting fetal condition was studied. The Apgar score of less than 7 was used as the judgment standard of neonatal asphyxia. The results showed that 18 cases (1.3%) of 1320 reactive types and 36 cases (45%) of 80 nonreactive types had neonatal asphyxia[7]. Huang, Q., and others proposed that fetal distress is when the fetal heart rate is more than 160 or less than 100 times per minute. Irregular fetal rate and fetal agitation are also manifestations of fetal hypoxia [8]. Johnston, J. C., and others believe that the application of Internet technology and short-range wireless transmission technology to the electronic fetal monitoring system has become a hot spot of research for current experts and scholars. The traditional fetal monitoring mode has also changed greatly, and the traditional mode of "arriving at the hospital after the disease" has begun to move toward the modern medical mode of "early prevention and early active diagnosis and treatment" [9]. Oweis, R. and others believe that the traditional fetal monitoring equipment can only be monitored in the hospital. The monitoring host is huge and bulky. Pregnant women have a baby and heavy body. For each prenatal examination, the patient needs to travel back and forth between home and the

hospital, which brings great trouble to pregnant women [10]. When Pitts, D. S. and others found the development and application of wireless communication technology and Internet technology. Various portable wireless monitoring devices appear, such as portable fetal heart rate monitor, central monitoring station, and intelligent electronic fetal monitoring system [11].

3. Experimental Analysis

3.1. Data and Methods

3.1.1. General Information. 230 pregnant women in maternal and child health care hospital were selected to form the experimental group. They were 21 ~ 39 years old, with an average of 26.6 ± 5.1 years old. The gestational weeks were 36 ~ 42 weeks, with an average of 39.8 weeks. Among them, 170 cases were primipara and 60 cases were postmenopausal women; 190 cases were normal pregnancy and 30 cases were high-risk pregnancy. 200 parturients in the hospital with detailed records in the past were selected as the control group; they were 21–41 years old, with a mean of (25.5 ± 5.6) years old; the gestation week lasts 35 to 42 weeks, with an average of 39.5 weeks; among them, 154 cases were primipara and 46 cases were postmenopausal women; 172 cases were normal pregnancy and 28 cases were high-risk pregnancy. There were no significant differences in age, gestational period, and total pregnancy rate between the two groups ($p > 0.05$). The details are shown in Table 1 below.

3.1.2. Method. When the pregnant woman is in the non-starvation state, the body position is taken as the semi-recumbent position or the left recumbent position. The ultrasonic Doppler probe is fixed at the position with the clearest fetal heart sound, and the uterine contraction probe is fixed at the two transverse fingers under the uterus to receive the signal. The fetus is monitored routinely for 20 minutes. In case of abnormality, it shall be extended to 30 ~ 60 min and rechecked in time to eliminate the impact of fetal sleep, pregnant women's spirit, and other factors on the monitoring results. If the uterine orifice of pregnant women expands more than 2 cm, continuous monitoring shall be carried out [12], with a minimum time of 0.5 h and a maximum time of 7.5 h, and no sedatives shall be used. Non-stimulation test, oxytocin test, and uterine contraction test were carried out according to relevant standards.

(1) NST Examination. All pregnant women had routine NST examination once a week from the 36th week of pregnancy until delivery. The pregnant woman takes the semi-recumbent position and the ultrasonic probe is placed on the abdomen (fetal heart sound area) after coating the coupling agent, and fixed with an abdominal band. The paper feeding speed is set at 3 cm/min. While tracing the fetal heart rate, when the pregnant woman feels that fetal movement occurs, the fetal movement tracing button is pressed by hand to make a mark on the paper [12]. Continuous recording shall not be less than 20 min. If the fetal heart rate is 120–160 beats/min, the baseline rate rises to 15 beats/min or more,

TABLE 1: Experimental object situation table.

	Age	Gestational week (weeks)	Primipara	Normal pregnancy	High-risk pregnancy
Experience group	26.6 ± 5.1	39.8	170	190	30
Control group	25.5 ± 5.6	39.8	154	46	172

and the fetal movement is 2–4 or more, and each time lasts for more than 15s, it is a reactive type; if the abovementioned indicators are not met, the monitoring time shall be extended for 30–40 min, and the examiner shall gently push the abdomen of the pregnant woman to stimulate the fetus. If the abovementioned indicators can be met, it is still considered that NST has a response [13]. For those who fail to meet the above indicators, the nonreactive type can be judged only after monitoring for at least 40 min, the baseline variability is below 6bpm, there is no fetal movement, or there is fetal movement without fetal heart rate acceleration and the duration of acceleration amplitude is below the reactive type standard.

3.1.3. Observation Index. Cases of fetal complications, neonatal asphyxia, and cesarean delivery have been reported. Diagnostic criteria of fetal distress: when the fetal heart rate is ≤ 120 beats/min or ≥160 beats/min, it indicates that the fetus has hypoxia; in terms of fetal movement, at first, fetal movement was frequent, and then fetal movement decreased or disappeared; amniotic fluid: those with grade II ~ III fecal contamination; and in terms of umbilical cord, the S/D ratio of umbilical cord blood flow ≥2.6. Diagnostic criteria of neonatal asphyxia: mild asphyxia with an Apgar score of 4 ~ 7 at 1 minute of birth; an Apgar score of 0 ~ 3 at 1 min of birth is severe asphyxia [12].

3.2. Result Analysis. According to statistics, the fetal complications and cesarean section rates in the experimental group were 16.38% (36/220) and 36.81% (82/220), which is significantly higher than 5.40% (9/200) and 17.60% (35/200) ($P < 0.05$) in the control group. The mild and severe asphyxia of newborns in the experimental group were observed 3.28% (7/220) and 1.43% (3/220), respectively, significantly lower than 9.60% (19/200)) and 6.50% (13/200) of the control group ($p < 0.05$). See Table 2.

Relationship between NST examination results and neonatal asphyxia: there were 190 cases of the NST reactive type, including 10 cases of mild neonatal asphyxia, accounting for 1.18%. There were 49 cases of the NST unresponsive type, including 18 cases of neonatal mild asphyxia, accounting for 36.73%, 2 cases of severe asphyxia, accounting for 4.09%, and 20 cases of asphyxia, accounting for 40.82%. There was a very significant difference between the two groups ($p < 0.01$). Figure 2 shows the specific experimental results.

4. Discussion

According to the existing research results, neonatal asphyxia is still a very important cause of neonatal death and disability. The latest foreign reports show that its incidence is 5% ~ 6%, while in China, it also has an incidence of 4.7% ~

8.9%. According to the survey results of 18 cities in China, asphyxia and its complications account for 33.5% of the death causes of newborns. It is generally believed that the occurrence of neonatal asphyxia and complications is a continuation of fetal intrauterine distress, which indicates that in the clinical process, we should study the causes of fetal intrauterine distress and deal with it timely and effectively. Fetal intrauterine distress generally refers to a series of clinical manifestations caused by fetal circulatory hypoxia in the uterus, which is an important complication during delivery; in terms of clinical manifestations, it is mainly the abnormal fetal heart rate and decreased or disappeared fetal movement. In the actual clinical work [14], researchers generally take the abnormality of fetal heart rate and meconium pollution as the main diagnostic criteria, this requires careful monitoring of changes in the fetal heart rate. This is caused by fetal hypoxia, the peristalsis of the intestine shows a certain hyperactivity, while the anal sphincter is relaxed, and the meconium is discharged into the amniotic fluid, causing pollution. The fetus inhales the meconium into the trachea, aggravating the hypoxia. This requires medical workers to clear the respiratory tract after the delivery of the fetus and ensure the smoothness of the respiratory tract, which has become the primary task of rescuing neonatal asphyxia. In the process of observation, if there is serious meconium pollution, they can choose to terminate the pregnancy immediately to prevent asphyxia and other complications. Various factors of blood gas exchange between the mother and fetus can cause fetal distress. Various factors of decreasing blood oxygen saturation can lead to neonatal asphyxia, which can occur at any stage of pregnancy, but most of them occur after the beginning of labor. If there is severe hypoxia in this process, it may cause fetal death in uterus [14].

Changes in fetal heartbeat are regulated by the fetal central nervous system, and fetal heartbeat acceleration and fetal movement are considered to be the circulatory system's response to fetal movement. It is believed that stimulating fetal movement causes the fetal heartbeat to accelerate through the central nervous system, such as the cerebral circulation. Therefore, fetal heart monitoring is the monitoring of the cardiac regulation function of the fetal central nervous system. In order to correctly respond to the health status of the fetal central nervous system, it can be used to judge whether the fetus has intrauterine hypoxia and its severity, and take effective measures according to specific conditions, such as intrauterine resuscitation or termination of pregnancy, which can prevent fetal acidosis and secondary damage to fetal heart, brain, liver, kidney and other tissues and cells; improve the prognosis of the fetus; improve the occurrence of various complications and sequelae of the newborn; and improve the quality of life of the newborn. In the process of monitoring, the midwife should closely

TABLE 2: Comparison of fetal complications and cesarean section between the two groups (N%).

Group	Number of cases	Fetal distress	Mild neonatal asphyxia	Severe neonatal asphyxia	Cesarean section
Experience group	230	36 (16.38)	7 (3.28)	3 (1.43)	82 (36.81)
Control group	200	9 (5.40)	20 (9.60)	13 (6.50)	34 (17.60)
p Value		<0.05	<0.05	<0.05	<0.05
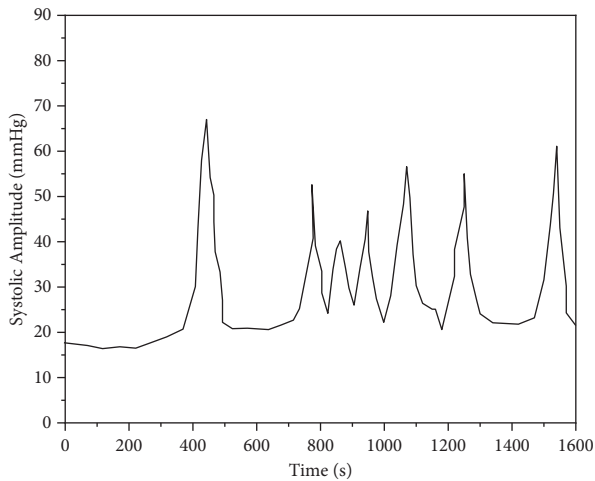
FIGURE 2: Relationship between NST examination results and neonatal asphyxia.

observe the changes of fetal heart rate and uterine contraction pressure, make various relevant necessary records, and adjust the probe position in time according to the situation, so that the fetal heart signal transmission is the clearest and most convenient, and the results obtained are the most authentic and credible. It also requires obstetricians to have a clear and full understanding of various changes of fetal heart rate, make sufficient preparations before fetal delivery, correctly deal with each stage of labor, and actively prevent the occurrence of premature birth and low birth weight infants, which play a positive role in reducing the occurrence of neonatal asphyxia.

4.1. Algorithm for Deriving the Fetal Heart Rate Curve

4.1.1. Analysis of the Fetal Heart Rate Baseline Extraction Algorithm. Baseline FHR is defined as the average fetal heart rate over 10 minutes in the absence of fetal movement or uterine contractions as the fetal heart rate base (e.g., FHR base). The fetal heart rate graph recorded on the monitor is a fluctuating band-like curve. According to the baseline fetal heart rate, the fetal heart rate curve can be divided into three types: tachycardia, normal, and bradycardia. The classification framework is shown in Table 3.

The fetal heart rate is susceptible to stimuli such as fetal movement, contractions, and palpation, and will temporarily accelerate and decelerate (for 10–20 s) before returning to the baseline. Because it is easy to update the fetal base and body time, it is necessary to accurately calculate the heart rate throughout the fetal period [15].

TABLE 3: Baseline classification of the fetal heart rate.

Classification		Variation range (times/minute)
Normal		120–160
Tachycardia	Light	161–180
	Severe	Above 180
Bradycardia	Light	100–119
	Severe	Below 99

Based on the clinical characteristics of the fetal heart rate, a method for estimating baseline fetal heart rate was developed in this paper. The main idea of the algorithm is as follows: (1) First read the fetal heart rate data, delete the fetal heart rate data (data other than 90–200 times/min), analyze the remaining data with histogram, and record the fetal heart rate and histogram with FHR, as the base value, accounts for the largest proportion of the remaining data. (2) Taking the baseline as the reference point, the fetal heart curve is smoothed and filtered. (3) The fetal heart rate curve varies greatly, and it is not recommended to use the smoothening algorithm alone. Therefore, the steep part of the change must be corrected and filtered after the threshold is set. (4) After four thresholds of correction and repeated processing, a relatively uniform baseline fetal heart rate can be obtained.

When obtaining the position of reference point through histogram analysis, first the fetal instantaneous heart rate from BPM to time domain unit (fetal cardiac cycle) is converted to facilitate the calculation and screening of the fetal heart rate data. For example, the selected fetal heart rate range is 90bpm-200bpm and the time domain is 300ms-600 ms. The fetal heart rate reference point can be reflected by finding a relatively concentrated fetal heart cycle in a sampling point through the histogram. As shown in Figure 3, the position and size of the reference point (inverted triangle identification) can be obtained by intuitive observation.

The smoothening filtering adopts forward and backward smoothening at the same time. The obtained reference point is used as the reference point B_0 for the smooth filtering of the fetal heart rate, and the filtering processing is carried out on the basis of B_0. The forward and backward filtering formulas are as follows:

Initial position: $B_0 = 0.975B_0 + 0.025B_i$
Forward filtering: $B_i = 0.965B_{i+1} + 0.035B_i$
Backward filtering: $B_i = 0.965B_{i-1} + 0.035B_i$

If the fetal heart rate fluctuates greatly, smoothening is not recommended, and the part with large fluctuation needs to be corrected. The process of correcting the algorithm is to preset an upper and lower threshold. When the fetal heart rate data are greater than the upper threshold of the baseline, the current fetal heart rate value is replaced. According to the
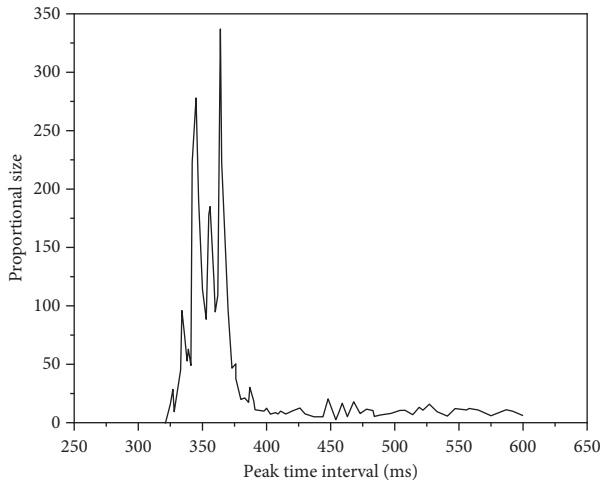
Figure 3: Analysis of the fetal heart rate baseline by a histogram method.

set threshold, the position and size of the point to be replaced can be found. Similarly, when the fetal heart rate data are less than the lower threshold of the flow baseline [16], it is replaced in the same way. The closed value parameter settings of three iterations are shown in Table 4.

As shown in Figure 4, the baseline of fetal heart rate during the whole monitoring process is obtained. The thin line with large fluctuation is the instantaneous rate curve of the fetal heart rate, and the thick line more stability and smoothness is the baseline of the fetal heart rate.

4.1.2. Analysis of FHR Acceleration and FHR Deceleration Algorithms.

Fetal heart rate acceleration is defined as a temporary increase in fetal heart rate at baseline, the fetus is not less than twice within 20 minutes after the beginning of the experiment, and the instantaneous rate is greater than the baseline of fetal heart rate by more than 15bpm for more than 15 s. In the process of fetal development, acceleration begins to appear about 18 weeks of pregnancy, and the improvement of acceleration mechanism is after 28–29 weeks. Therefore, acceleration is an important physiological phenomenon in the third trimester of pregnancy and an index to judge the health and safety of the fetus in the womb. There are two main types of acceleration: periodic acceleration and aperiodic acceleration. Non-periodic acceleration occurs with fetal movement, internal diagnosis, or abdominal palpation. Periodic contractions accompany uterine contractions, with an increase in the fetal heart rate occurring in tandem with the contractions. Based on the fetal heart rate definition, an algorithm was developed to accelerate the fetal heart rate. The basic process is as follows: (1) obtain the baseline of fetal heart rate and detect the instantaneous FHR value of each sampling point. (2) Find the data greater than 10bpm of FHR baseline value, search the maximum value in the next minute, and record the peak value and occurrence time. (3) Then, search for points 3 BPM larger than the baseline value in 55s forward and backward respectively. (4) In the third step, search the

Table 4: Parameter setting of baseline correction threshold of the fetal heart rate.

Number of iterations	Upper threshold (bpm)	Lower threshold (bpm)
1	25	20
2	20	15
3	10	10
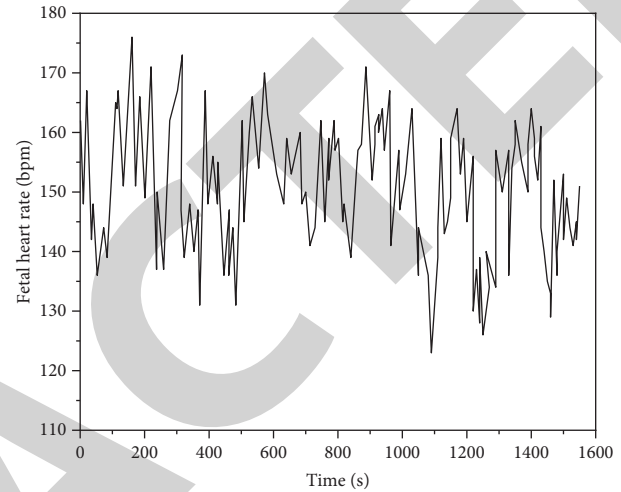
Figure 4: Baseline detection and analysis of fetal heart rate.

FHR minimum value point of the corresponding time range, and record the occurrence time as the starting point and ending point [17].

The definition of fetal heart rate deceleration is the temporary deceleration of fetal heart rate accompanied by uterine contraction. It can be divided into periodic deceleration and aperiodic deceleration according to the occurrence time, shape, and continuous length. At present, most of them still use Edward Hon's method to classify, record the fetal heart rate and uterine contraction curve for 40–60 minutes, and then see whether the deceleration pattern of each fetal heart rate is basically the same and whether the deceleration waveform has a certain relationship with the uterine contraction waveform. Periodic deceleration is judged according to the time-position relationship between the fetal heart rate curve and uterine contraction. It is divided into early deceleration, late deceleration, and variable deceleration. Their clinical manifestations are as follows: (1) Early deceleration generally occurs in the later stage of the first stage of labor. Uterine contraction is caused by fetal head compression and is not changed by the pregnant women's body position or oxygen inhalation. (2) Mutation decelerates, suggesting umbilical cord compression. (3) The classification range of late deceleration, placental dysfunction, and fetal hypoxia is shown in Table 5.

The idea and process of deceleration detection algorithm is similar to that of acceleration algorithm: (1) obtain the baseline of fetal heart rate and detect the instantaneous FHR value of each sampling point. (2) Find the data less than 20bpm of FHR baseline value, and search the minimum

TABLE 5: Types of fetal heart rate deceleration.

	Decrease of fetal heart rate	Lag time	Deceleration duration
Get up early and slow down	[l5, 50] bpm	≤15 s	60–70 s
Late deceleration	[15, 30] bpm	[30, 60] s	60–70 s
Mutation deceleration	≥15 bpm	Uncertain	15–20 s
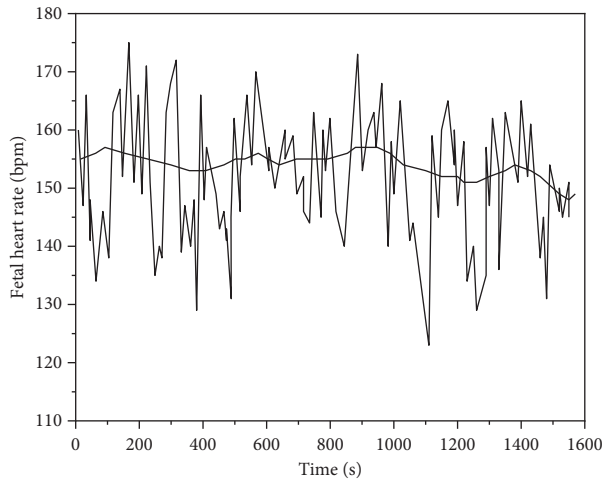
FIGURE 5: Analysis of fetal heart rate deceleration and deceleration detection.

value, marked trough value and occurrence time in the next minute. (3) Then, search for points 3 BPM smaller than the baseline value in 55s forward and backward, respectively. (4) Find the point that meets the conditions (FHR in the corresponding range is the fetal heart rate deceleration event) and mark the decline amplitude and occurrence time. The actual detection effect is shown in Figure 5. The acceleration event points detected above the baseline are marked with "∗", and the deceleration event points detected below the baseline are marked with "+", so that the number and time of fetal heart rate acceleration and deceleration can be obtained.

5. Conclusion

In this study, the incidence of fetal distress and cesarean section rate in the experimental group were significantly higher than those in the control group ($p < 0.05$). The rates of mild and severe neonatal asphyxia in the experimental group were significantly lower than those in the control group ($p < 0.05$). To sum up, the correct application of fetal heart rate monitoring in the perinatal period can early detect and deal with fetal distress and reduce the occurrence of various complications such as neonatal asphyxia, which is worthy of popularization and application in clinics. The smart electronic fetal monitoring system realizes intelligent monitoring through Internet of things technology, Internet platform, and portable sensor equipment, and uploads the monitoring data to the cloud. Doctors and pregnant women interact through the intelligent maternal and infant care platform based on the Internet of things and cloud platform. Doctors can timely understand the health status of pregnant women and fetuses and remotely guide them to carry out corresponding operations. Pregnant women can enjoy the same supervision as in the hospital anytime and anywhere. The whole maternal and infant monitoring system constitutes a network system with monitoring, analysis, and feedback functions, which realizes smart monitoring anytime and anywhere, and reduces the pain of patients running between families and hospitals. In terms of fetal heart rate detection, obstetric experts and large sample clinical data need to be tested repeatedly to optimize and improve the recognition algorithm of relevant parameters of fetal heart rate contraction curve, so as to improve its accuracy and reliability.

Data Availability

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

Ethical Approval

All procedures performed in studies involving human participants were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. Wang, "Study of the relationship in pregnant women between hepatitis b markers and a placenta positive for hepatitis b surface antigen: journal of perinatal medicine," *Journal of Perinatal Medicine*, vol. 43, no. 2, pp. 191–199, 2015.

[2] E. Lim, H.-K. Lee, H.-S. Myoung, and K.-J. Lee, "Development of a noncontact heart rate monitoring system for sedentary behavior based on an accelerometer attached to a chair," *Physiological Measurement*, vol. 36, no. 3, pp. N61–N70, 2015.

[3] Z. Jia, Z. Li, Z. Yang, G. Ren, and J. Lin, "Design and implementation of a coal-bed methane fracture monitoring system based on virtual instrument technology," *Instrumentation Science & Technology*, vol. 43, no. 3, pp. 290–302, 2015.

[4] O. Gomez, C. Portilla, and M. A. Rios, "Reliability analysis of substation monitoring systems based on branch pmus," *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 962–969, 2015.

[5] J. Han, J.-D. Jeong, I. Lee, and S.-H. Kim, "Low-cost monitoring of photovoltaic systems at panel level in residential homes based on power line communication," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 4, pp. 435–441, 2017.

Hindawi

Research Article

A Flexible Bayesian Parametric Proportional Hazard Model: Simulation and Applications to Right-Censored Healthcare Data

Abdisalam Hassan Muse ⓘ,[1] **Oscar Ngesa,**[2] **Samuel Mwalili** ⓘ,[3] **Huda M. Alshanbari** ⓘ,[4] **and Abdal-Aziz H. El-Bagoury** ⓘ[5]

[1]*Department of Mathematics (Statistics Option),*
 Pan African University Institute for Basic Science Technology and Innovation (PAUSTI), Nairobi 62000 00200, Kenya
[2]*Department of Mathematics and Physical Sciences, Taita Taveta University, Voi 635-80300, Kenya*
[3]*Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology (JKUAT),*
 Nairobi, Kenya
[4]*Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428,*
 Riyadh 11671, Saudi Arabia
[5]*Basic Science Department, Higher Institute of Engineering and Technology, El-Mahala El-Kubra, Egypt*

Correspondence should be addressed to Abdisalam Hassan Muse; abdisalam.hassan@amoud.edu.so

Received 2 February 2022; Revised 24 March 2022; Accepted 18 April 2022; Published 2 June 2022

Academic Editor: Mohamed Elhoseny

Survival analysis is a collection of statistical techniques which examine the time it takes for an event to occur, and it is one of the most important fields in biomedical sciences and other variety of scientific disciplines. Furthermore, the computational rapid advancements in recent decades have advocated the application of Bayesian techniques in this field, giving a powerful and flexible alternative to the classical inference. The aim of this study is to consider the Bayesian inference for the generalized log-logistic proportional hazard model with applications to right-censored healthcare data sets. We assume an independent gamma prior for the baseline hazard parameters and a normal prior is placed on the regression coefficients. We then obtain the exact form of the joint posterior distribution of the regression coefficients and distributional parameters. The Bayesian estimates of the parameters of the proposed model are obtained using the Markov chain Monte Carlo (McMC) simulation technique. All computations are performed in Bayesian analysis using Gibbs sampling (BUGS) syntax that can be run with Just Another Gibbs Sampling (JAGS) from the R software. A detailed simulation study was used to assess the performance of the proposed parametric proportional hazard model. Two real-survival data problems in the healthcare are analyzed for illustration of the proposed model and for model comparison. Furthermore, the convergence diagnostic tests are presented and analyzed. Finally, our research found that the proposed parametric proportional hazard model performs well and could be beneficial in analyzing various types of survival data.

1. Introduction

The healthcare domain has evolved significantly in recent years as a result of computational developments. The use of Bayesian statistics in healthcare has encouraged the application of computational developments, providing a powerful and versatile alternative to traditional methodologies used in healthcare [1]. The progress of Bayesian approaches in healthcare aims to make an individual's life more affordable

and comfortable, similar to how smartphones have made life easier [2]. Despite the fact that the idea of applying computational Bayesian statistics to survival analysis dates back to the 19th century, McMC techniques are now garnering more attention in the literature because of abundant and cheap computation [3]. The application of deep learning to the context of parametric survival models was discussed by [4]. Through an efficient training process, the quality of the developed system improves. Data portioning is done three

times to confirm the trained algorithms (training-testing-validation) [5]. The main goal of this article is to present the Bayesian parametric proportional hazard model using BUGs syntax.

The statistical analysis of survival data is an essential topic in many fields, including medicine, biology, environmental science, healthcare, economics, engineering, social science, and epidemiology, among others. Probability distributions serve as the foundation for survival models. The family of distributions can be parametric, semiparametric, or nonparametric. The parametric survival models lead to more efficient and smaller standard errors of the estimates than semiparametric and nonparametric models [6] if the distributional assumption is correct, to be more specific.

In analyzing survival data, parametric survival models are crucial. The benefits of using parametric survival models include the following: (1) handling all types of censored data (left, right, interval, double, and middle); (2) application of survival analysis in a healthcare care problem and (3) producing better estimation when you have a theoretical expectation of the baseline hazard; also, (4) they can apply random effects—frailty models—and can also be used to estimate expected lives, not only hazard ratios like the accelerated failure time models [7].

The proportional hazards (PH) model, in which covariates affect the hazard rate function, and the accelerated failure time (AFT) model, in which covariates affect both the hazard rate and time scale, are the two most common methods for developing parametric regression models for survival data [8]. However, other class of models have also been proposed such as the accelerated hazard (AH) model [9] and the proportional odds (PO) model [10].

One of the first steps in using a parametric approach to model survival data is to choose a suitable baseline distribution that can capture significant features of the observations of interest. Certain probability distributions are widely used in the modelling of survival data. Only a few are closed under the proportional hazard model, and none are flexible enough to describe a wide range of survival data [11]. Most of the distributions closed under the PH assumption fails to model a nonmonotone (i.e., bathtub and unimodal) survival data sets.

The log-logistic (LL) distribution has a wide range of applications in survival data analysis and can accommodate unimodal survival data sets. The distribution is closed under both proportionality odds (PO) and multiplication of failure time (AFT) frameworks [7]. It is not a PH model, but an AFT model. However, when the log-logistic distribution is generalized, it has the appealing feature of being a member of all classes of parametric hazard-based regression models of the survival analysis because its failure rate function is quite versatile and its cumulative hazard function (chf) has a tractable form.

Extensive efforts have been made over the last decades to extend classical distributions to use as a baseline distribution for parametric hazard-based regression models. Many modifications to the LL distribution have been introduced to make it more adaptable to a wide range of hazard shapes

[12–16]. The generalized log-logistic distribution (GLL) is one such model, which modifies the log-logistic distribution by inducing an additional shape parameter [17]. The model is tractable and closed under the PH assumption and can account for both nonmonotone and monotone hazard rates [11]. On the other hand, recent computational advances have advocated for the use of Bayesian techniques in the field of survival and reliability analysis.

The motivating ideas behind our work on Bayesian parametric proportional hazard (PH) model with GLL baseline hazard are as follows: (i) despite the fact that there are some classical distributions closed under the PH framework, none of which is flexible enough to incorporate both monotone and nonmonotone hazard rate; (ii) Bayesian inference does not rely on asymptotic approximation for statistical inference; (iii) the availability of software makes Bayesian implementation for hazard-based complicated models relatively more straightforward and simple than classical inference [18]; (iv) parametric PH model may lead to more precise estimates than the semiparametric PH model; and, last but not the least, (v) the use of generalized distributions that can capture both monotone and nonmonotone hazard rate functions is what makes our work unique and more appealing to biostatisticians, epidemiologists, healthcare workers, and other applied researchers in multiple disciplines.

To the best of author's knowledge, no Bayesian inferences study has been conducted on the PH model with generalized log-logistic baseline hazard. As a result, in this paper, we consider the Bayesian inference for the generalized log-logistic proportional hazard model, beginning with the PH model formulation and assumptions, revising the generalized log-logistic distribution, and verifying that the GLL distribution is closed under the PH framework. In addition, we discuss the inferential procedures and how to obtain the classical and Bayesian estimators for the model's parameters. We also compare the proposed model to other existing distributions closed under the PH framework, and one interesting feature of this model is that it can incorporate different hazard rate shapes. Hence, the formulation of the parametric PH model and its lifetime function, the inferential procedures using both classical and Bayesian approaches, and the development of the computational algorithms to fit the proposed PH model and its competing models using RJAGS in R software are the novelty of this study.

The article is structured as follows: the PH model formulation, assumptions, and its probabilistic functions are discussed in Section 2. Section 3 revises the most common probability distributions closed under the PH model. Section 4 presents the proposed baseline hazard function which is a generalized log-logistic (GLL) distribution. The GLL distribution under the PH model is presented in Section 5. Section 6 discusses the inferential procedures of the proposed model. In Section 7, we present an McMC simulation study to assess the performance of the proposed model. Section 8 presents the application of the proposed model to two right-censored cancer data sets with monotone and nonmontone hazard rates. In addition, the convergence

diagnostics of the McMC techniques were discussed. The Bayesian model selection criterion is presented in Section 9. Finally, in the final portion, the article's concluding notes are offered, and some future works are mentioned.

2. PH Model Formulation and Assumptions

In many real-life applications, survival times are affected by explanatory variables. The explanatory variable vector is related to response variable through a regression model. An important aspect of survival modelling is the inclusion of explanatory variables. The hazard-based regression models can be formulated in a number of ways. One of the most frequently used method is the proportional hazard (PH) model formulation.

PH models play an essential role in analyzing time-to-event data and are broadly used in survival and reliability analysis as well as in joint modelling of survival and longitudinal data [7]. It is the most popular parametric model in medical studies and clinical trials because of the existence of a semiparametric PH hazard model which is robust against the distributional assumption of the survival time. The parametric PH model is given with the similar form to the Cox PH model. It is the parametric form of the Cox PH models [6].

2.1. PH Model Formulation.
The parametric proportional hazard (PH) models are formulated using a defined baseline hazard and a link function $\psi(\mathbf{x}'\beta)$ for the covariates which is defined as follows:

$$
\begin{aligned}
&\psi(\mathbf{x}'\beta) > 0, \ \forall x \neq 0, \\
&\quad \psi(\mathbf{x}'\beta) \text{ is a monotone function that has a one} \\
&\qquad \text{-to-one correspondence,} \\
&\quad \psi(0) = 1.
\end{aligned}
\tag{1}
$$

The most commonly found option for the link function $\psi(\mathbf{x}'\beta)$ is the exponential $\exp(\mathbf{x}'\beta)$ (or log-linear) function. In this work, we define the PH model with the assumption that $\psi(\mathbf{x}'\beta) = \exp(\mathbf{x}'\beta)$.

2.2. PH Assumptions.
The PH model assumption is that the effect of covariates is to increase or decrease the hazard rate function by a proportionate amount which does not depend on t. The assumption of the PH model can be defined as follows:

$$
h(t; x) = h_0(t)\psi(\mathbf{x}'\beta) = h_0(t)\exp(\mathbf{x}'\beta) = h_0(t)e^{x'\beta}, \tag{2}
$$

where $h_0(t)$ is called the baseline hazard.

Simplifying, we get,

$$
h(t|x) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p). \tag{3}
$$

The main difference between the Cox PH model and the parametric PH model is that the baseline hazard function is assumed to follow a specific distribution when it is fitted to the data. Using equation (2), we can see that the hazard ratio

(HR) comparing any two specifications of the covariates, for example, $(\mathbf{x} \text{ and } \mathbf{x}^*)$ is

$$
\begin{aligned}
HR(\mathbf{x}, \mathbf{x}*, h_0, \beta) &= \frac{h(t|\mathbf{x}, \boldsymbol{\beta})}{h(t|\mathbf{x}^*, \boldsymbol{\beta})} = \frac{h_0(t|\mathbf{x})\exp(\boldsymbol{\beta}\mathbf{x}')}{h_0(t|\mathbf{x})\exp(\boldsymbol{\beta}\mathbf{x}^*)} \\
&= \exp\left[\left(\mathbf{x}' - \mathbf{x}^{*'}\right)^T \boldsymbol{\beta}\right].
\end{aligned}
\tag{4}
$$

The above equation shows us that the baseline hazards cancel each other from this ratio, so the hazard rate for one individual is proportional to the hazard rate for any other individual. On the other hand, the proportionality constant is independent of time which makes the main assumption of this model [6]. As a result, the model is known as the proportional hazard (PH) model in the literature.

Unlike most parametric regression models including accelerated failure time (AFT) models, PH models does not include an intercept [19]. More properly, the vector $\mathbf{X}$ in the PH model is not assumed to have $x \equiv 1$. An intercept would get confounded with the baseline hazard function h_0.

2.3. Lifetime Functions Describing the PH Model.
The five frequent representatives of a lifetime distribution function that are used to characterize the PH model are addressed in this section.

2.3.1. Hazard Rate Function of the PH Model.
In the PH analysis, one of the most important lifetime functions is the concept of the hazard rate function (hrf). The hazard rate function $h(t|x)$, abbreviated by hrf, also called the instantaneous failure rate or as force of mortality of a PH model is of the form:

$$
h(t; x) = h_0(t)\psi(\mathbf{x}'\beta) = h_0(t)\exp(\mathbf{x}'\beta) = h_0(t)e^{x'\beta}. \tag{5}
$$

2.3.2. Cumulative Hazard Function of the PH Model.
The hazard or survival functions, rather than the cumulative distribution or probability density function, are typically used in the PH analysis of survival data. The hazard rate function is used to interpret the most common survival regression models; however, the cumulative hazard function (chf), also known as the integrated hazard rate function, can be easily written down. Hence, the chf of a PH model takes the following form:

$$
H(t|x) = \int_0^t h(s; x)\,ds = e^{x'\beta} \int_0^t h_0(s) \, ds = e^{x'\beta} H_0(t). \tag{6}
$$

2.3.3. Survival Function of the PH Model.
The survivor function (sf) for a PH model can be derived using the following relationship between survival function and the hazard rate function. Hazard function is given by

$$h(t|x) = \frac{f(t|x)}{S(t|x)}. \tag{7}$$

Cumulative hazard function:

$$H(t|x) = \int_0^t h(u)\mathrm{du} = \int_0^t \frac{f(u)}{S(u)}\mathrm{du} = \int_0^t \frac{-\mathrm{d}S(u)}{S(u)}\mathrm{du}$$

$$= -\log\{S(t)\}, \tag{8}$$

$$f(t|x) = h(t|x)S(t|x) = h(t|x)\exp\{-H(t|x)\}.$$

Using the above expressions, we can easily find

$$S(t|x) = \exp\{-H(t|x)\},$$

$$S(t|x) = \exp\{-H(t|x)\} = \exp\left\{-\int_0^t \psi(x)h_0(t)\,dt\right\},$$

$$= \exp\left\{-\psi(x)\int_0^t h_0(t)\,dt\right\}, \tag{9}$$

$$= \left[\exp\left\{-\int_0^t h_0(t)\,dt\right\}\right]^{\psi(x)},$$

$$= [S_0(t)]^{\psi(x)}.$$

2.3.4. Cumulative Distribution Function of the PH Model.
The cdf of the PH model, also known as the lifetime distribution function, is given by

$$F(t) = 1 - S(t) = 1 - \exp\{-H(t)\},$$

$$F(t) = 1 - [S_0(t)]^{\psi(x)}. \tag{10}$$

2.3.5. Probability Density Function of the PH Model.
The pdf or the failure density function of the PH model is defined as

$$f(t) = f_0(t)\psi(x)[S_0(t)]^{\psi(x)-1}. \tag{11}$$

The five representatives used here were chosen for their special meaning for lifetime data, their intuitive appeal, their utility in survival data analysis, and, last but not the least, their popularity in probability theory and statistics.

The PH model can be formulated without assuming a probability distribution for survival times, and this leads to the well-known Cox PH model [20]. On the other hand, assuming a probability distribution for survival times leads to the fully parametric PH model. The most common parametric survival models used are as follows: exponential, Weibull, Gompertz, log-logistic, log-normal, gamma, and the generalized gamma distributions. Only the exponential, Weibull, and Gompertz distributions are used for the PH model. The log-logistic and the log-normal distributions are not closed under the PH framework. Weibull distribution is the only one that is closed under both parametric AFT and PH models.

3. Distributions Closed under PH Framework

In this section, we present most common parametric distributions that are closed under the PH framework and are used to analyze survival data. These distributions have been studied and used in various contexts in the literature.

3.1. Exponential PH Model.
Exponential distribution is a continuous probability distribution with only one unknown parameter k. It is the simplest distribution for lifetime distribution models. The distribution is not flexible enough to describe commonly encountered hazard rate shapes for survival data. The pdf, cdf, sf, hrf, and chf of the exponential random variable are, respectively, as follows.

Let $X \sim$ exponential (k),

$$f(t) = k \exp\{-kt\},$$

$$F(t) = 1 - \exp\{-kt\},$$

$$S(t) = \exp\{-kt\}, \tag{12}$$

$$h(t) = k,$$

$$H(t) = -\log S(t) = -\log(\exp\{-kt\} = kt,$$

where $k > 0$ is the scale parameter and $t \geq 0$. A short value of k shows low risk and long survival, where a large value shows high risk and short survival. For the PH model, the exponential baseline hazard is

$$h(t) = k. \tag{13}$$

So, according to the formulation of the PH framework, the hazard rate for an individual with covariate vector x and link function $\psi(x)$ is

$$h(t) = h_0(t)\,\psi(x) = k.\psi(x). \tag{14}$$

Applying the log-linear function $\psi(\mathbf{x}'\beta) = \exp(\mathbf{x}'\beta)$, we can simplify into

$$h_{EPH}(t) = k.\exp(\mathbf{x}'\beta) = k.\exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p). \tag{15}$$

In this equation, the hrf has the exponential distribution with scale parameter k. $\exp(\mathbf{x}'\beta)$ which indicates that the PH assumption is satisfied with the exponential distribution. It is worth mentioning that the exponential distribution is often found to be inadequate to describe survival data. This makes the applicability of this distribution fairly limited.

The other lifetime distributions of the exponential PH model are as follows.

The survival function of the exponential PH model is

$$S_{EPH}(t) = [\exp\{-kt\}]^{\exp(\mathbf{x}'\beta)}. \tag{16}$$

The pdf of the exponential PH model is

$$f_{EPH}(t) = k\exp\{-kt\}\exp(\mathbf{x}'\beta)[\exp\{-kt\}]^{\exp(\mathbf{x}'\beta)-1}. \tag{17}$$

The cdf of the exponential PH model is

$$F_{EPH}(t) = 1 - [\exp\{-kt\}]^{\exp(\mathbf{x}'\beta)}. \tag{18}$$

The chf of the exponential PH model is

$$H_{\mathrm{EPH}}(t) = kt \, \exp(\mathbf{x}'\beta). \tag{19}$$

3.2. Gompertz PH Model.

Gompertz distribution is named after Benjamin Gompertz, a British mathematician and actuary, who developed it in 1825. It is a continuous probability distribution used for modelling adult life spans and other application under different disciplines such as actuarial science, demography, survival, and reliability analysis. This distribution is flexible and can be skewed both in right and in left. The pdf, cdf, sf, hrf, and chf of the exponential random variable are, respectively, as follows.

Let $X \sim \mathrm{Gompertz}(k, \alpha)$,

$$
\begin{aligned}
f(t) &= \alpha k.e^{tk} \exp\{-\alpha(e^{tk} - 1\}, \, t \in [0, \infty), \\
F(t) &= 1 - \exp\{-\alpha(e^{tk} - 1\}, \\
S(t) &= \exp\{-\alpha(e^{tk} - 1\}, \\
h(t) &= = \alpha k e^{tk}, \\
H(t) &= -\log S(t) = -\log\left[\exp\{-\alpha(e^{tk} - 1\}\right] = \alpha(e^{tk} - 1),
\end{aligned}
\tag{20}
$$

where $k > 0$ is the rate parameter, $\alpha > 0$ is the shape parameter, and $t \geq 0$. When $k \geq 0$, the survival time then has an exponential distribution; therefore, Gompertz distribution is a generalization of exponential distribution. For the PH model, the Gompertz baseline hazard rate function is given by

$$h(t) = \alpha k e^{tk}. \tag{21}$$

So, according to the formulation of the PH framework, the hazard rate for an individual with covariate vector x and link function $\psi(x)$ is

$$h(t) = h_0(t)\,\psi(x) = \alpha k.e^{tk}.\psi(x). \tag{22}$$

Applying the log-linear function $\psi(\mathbf{x}'\beta) = \exp(\mathbf{x}'\beta)$, we can simplify into

$$h_{\mathrm{GoPH}}(t) = \alpha k.e^{tk}.\exp(\mathbf{x}'\beta) = \alpha k.e^{tk}.\exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p). \tag{23}$$

In the above equation, it is straightforward that the PH property is satisfied. However, the Gompertz PH model is rarely used in the real-life applications.

The other lifetime distributions of the Gompertz PH model are as follows: the survival function of the Gompertz PH model is

$$S_{\mathrm{GoPH}}(t) = \left[\exp\{-\alpha(e^{tk} - 1\}\right]^{\exp(\mathbf{x}'\beta)}. \tag{24}$$

The pdf of the Gompertz PH model is

$$f_{\mathrm{GoPH}}(t) = \alpha k.e^{tk} \exp\{-\alpha(e^{tk} - 1\}\exp(\mathbf{x}'\beta)\left[\exp\{-\alpha(e^{tk} - 1\}\right]^{\exp(\mathbf{x}'\beta)-1}. \tag{25}$$

The cdf of the Gompertz PH model is

$$F_{\mathrm{GoPH}}(t) = 1 - \left[\exp\{-\alpha(e^{tk} - 1\}\right]^{\exp(\mathbf{x}'\beta)}. \tag{26}$$

The chf of the Gompertz PH model is

$$H_{\mathrm{GoPH}}(t) = \exp(\mathbf{x}'\beta)\,\alpha(e^{tk} - 1). \tag{27}$$

3.3. Weibull PH Model.

Weibull distribution is a generalization of the exponential distribution. It is a versatile distribution that can take on the characteristics of other types of continuous distributions. It has an additional parameter compared to the exponential. The additional parameter describes the shape of the hazard functions, based on the value of the shape parameter [21]. The pdf, cdf, sf, hrf, and chf of the Weibull random variable are, respectively, as follows.

Let $X \sim \mathrm{Weibull}(k, \alpha)$,

$$
\begin{aligned}
f(t) &= \alpha k(kt)^{\alpha-1} \exp\{-(kt)^{\alpha}\}, \\
F(t) &= 1 - \exp\{-(kt)^{\alpha}\}, \\
S(t) &= \exp\{-(kt)^{\alpha}\}, \\
h(t) &= \alpha k(kt)^{\alpha-1}, \\
H(t) &= -\log S(t) = -\log(\exp\{-(kt)^{\alpha}\} = (kt)^{\alpha},
\end{aligned}
\tag{28}
$$

where $\alpha > 0$ is the shape parameter and $k > 0$ is the rate parameter. The hazard rate function increases when $\alpha > 1$, decreases for $\alpha < 1$, and constant for $\alpha = 1$. When $\alpha = 1$, the Weibull distribution reduces to exponential. It is worth mentioning that the Weibull distribution does not accommodate nonmonotone (i.e., unimodal or bathtub) hazard rates.

For the PH model the Weibull baseline hazard is

$$h(t) = \alpha k(kt)^{\alpha-1}. \tag{29}$$

So, according to the formulation of the PH framework, the hazard rate for an individual with covariate vector x and link function $\psi(x)$ is

$$h(t) = h_0(t)\,\psi(x) = \alpha k(kt)^{\alpha-1}\psi(x). \tag{30}$$

Applying the log-linear function $\psi(\mathbf{x}'\beta) = \exp(\mathbf{x}'\beta)$, we can simplify into

$$
\begin{aligned}
h_{\mathrm{WPH}}(t) &= \alpha k(kt)^{\alpha-1} \exp(\mathbf{x}'\beta) \\
&= \alpha k(kt)^{\alpha-1} \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p).
\end{aligned}
\tag{31}
$$

In this equation, the model has the Weibull distribution with rate parameter $k. \exp(\mathbf{x}'\beta)$ and shape parameter α which indicates that the PH assumption is satisfied with the Weibull distribution with constant α.

The other lifetime distributions of the PH Weibull model are as follows: the survival function of the Weibull PH model is

$$S_{\mathrm{WPH}}(t) = \left[\exp\{-(kt)^{\alpha}\}\right]^{\exp(\mathbf{x}'\beta)}. \tag{32}$$

The pdf of the Weibull PH model is

$$f_{WPH}(t) = \alpha k(kt)^{\alpha-1} \exp\{-(kt)^{\alpha}\}\exp(\mathbf{x}'\beta)\left[\exp\{-(kt)^{\alpha}\}\right]^{\exp(\mathbf{x}'\beta)-1}. \tag{33}$$

The cdf of the Weibull PH model is

$$F_{\text{WPH}}(t) = 1 - \left[\exp\{-(kt)^{\alpha}\}\right]^{\exp(\mathbf{x}'\beta)}. \tag{34}$$

The chf of the Weibull PH model is

$$H_{\text{WPH}}(t) = \exp(\mathbf{x}'\beta)(kt)^{\alpha}. \tag{35}$$

4. Parametric Baseline Hazard

The parametric baseline hazard function is essential because it determines which hazard shapes can be captured by the proportional hazard (PH) model. Most classical distributions that are closed under the PH framework, such as the exponential, Weibull, and Gompertz distributions, are incapable of accommodating unimodal hazard shapes. As a result, it is worth looking into some modifications to the classical distributions that can account for both monotone and nonmonotone hazard rates.

In this paper, we consider the Bayesian inference for the parametric PH models with generalized log-logistic (GLL) baseline. The GLL is a flexible survival distribution proposed by [11]. This distribution has a characteristic similar to those of the log-logistic distribution. Also, the advantage of the GLL distribution is that it approaches to Weibull in the limit. These properties allowed the GLL to handle both monotone and nonmonotone hazard functions, and also it makes to be a baseline distribution that is closed under both AFT and PH model [22] like the Weibull distribution. The distribution is adaptable, and the two shape parameters enable a wide range of hazard shapes. It also includes a variety of important distributions such as the exponential, Weibull, Burr XII, and log-logistic distributions. In addition, when compared to competitors, it is relatively tractable. We refer to, for more information on the distribution and its properties, [17].

For a positive-valued random variable T, the hrf of the GLL distribution with three unknown parameters $k > 0, \alpha > 0, \eta > 0$ is given by

$$h(t; \boldsymbol{\theta}) = \frac{\alpha k (kt)^{\alpha-1}}{[1 + (\eta t)^{\alpha}]} = \frac{\alpha k^{\alpha} t^{\alpha-1}}{[1 + (\eta t)^{\alpha}]}, \quad t \geq 0, \ k, \alpha, \eta > 0. \tag{36}$$

The chf of the GLL distribution is given by

$$H(t; \boldsymbol{\theta}) = \frac{k^{\alpha}}{\eta^{\alpha}} \log\left[1 + (\eta t)^{\alpha}\right], \quad t \geq 0, \ k, \alpha, \eta > 0. \tag{37}$$

The distribution function of the GLL model is of the form:

$$F(t; \boldsymbol{\theta}) = 1 - \left[1 + (\eta t)^{\alpha}\right]^{-k^{\alpha}/\eta^{\alpha}}, \quad t \geq 0, \ k, \alpha, \eta > 0. \tag{38}$$

The survival function (sf) of the GLL model is given by

$$S(t; \boldsymbol{\theta}) = \left[1 + (\eta t)^{\alpha}\right]^{-k^{\alpha}/\eta^{\alpha}}, \quad t \geq 0, \ k, \alpha, \eta > 0, \tag{39}$$

where $k > 0, \alpha > 0$, and $\eta > 0$ are parameters and $\boldsymbol{\theta} = (k, \alpha, \eta)'$.

The quantile function of the GLL model is given by

$$X_{q=} F^{-1}(q; k, \alpha, \eta) = \frac{\left\{[1/1 - q - 1]^{\eta^{\alpha}/k^{\alpha}} - 1\right\}^{1/\alpha}}{\eta}, \quad 0 \leq q < 1. \tag{40}$$

The reverse cumulative hazard rate function is expressed as follows:

$$H_0^{-1}(u; \boldsymbol{\theta}) = \frac{\left(e^{\eta^{\alpha} k^{-\alpha} u} - 1\right)^{1/\alpha}}{\eta}. \tag{41}$$

Figure 1 illustrates shapes that the failure rate function can accept such as constant, increasing, decreasing, V-shape, and unimodal among others.

5. The Proposed PH Model

For the PH model, the generalized log-logistic baseline hazard is

$$h(t) = \frac{\alpha k (kt)^{\alpha-1}}{[1 + (\eta t)^{\alpha}]}. \tag{42}$$

So, according to (2), the hazard rate for an individual with covariate vector x and link function $\psi(x)$ is

$$h(t) = h_0(t)\psi(x) = \frac{\alpha k (kt)^{\alpha-1}}{[1 + (\eta t)^{\alpha}]}\psi(x). \tag{43}$$

Applying the log-linear function $\psi(\mathbf{x}'\beta) = \exp(\mathbf{x}'\beta)$, we can simplify into

$$h_{\text{GLLPH}}(t) = \frac{\alpha k (kt)^{\alpha-1}}{[1 + (\eta t)^{\alpha}]}\exp(\mathbf{x}'\beta) = \frac{\alpha k^{\alpha} t^{\alpha-1}}{[1 + (\eta t)^{\alpha}]}\exp(\mathbf{x}'\beta)$$

$$= \frac{\alpha\left(k.\exp(\mathbf{x}'\beta)^{1/\alpha}\right)^{\alpha} t^{\alpha-1}}{[1 + (\eta t)^{\alpha}]} = \frac{\alpha k^{*\alpha} t^{\alpha-1}}{[1 + (\eta t)^{\alpha}]}. \tag{44}$$

In this equation, the hrf can be recognized as a generalized log-logistic distribution as well, but contrary to (36), the rate parameter is $k^* = k.\exp(\mathbf{x}'\beta)^{1/\alpha}$ and shape parameters are α and η which indicates that the PH assumption is satisfied with the GLL distribution and the proposed model is closed under the PH framework.

The other lifetime distribution functions for the GLL PH model are as follows: the survivor function of the GLL PH model is

$$S_{\text{GLLPH}}(t) = \left[\left[1 + (\eta t)^{\alpha}\right]^{-k^{\alpha}/\eta^{\alpha}}\right]^{\exp(\mathbf{x}'\beta)}. \tag{45}$$

The pdf of the GLL PH model is

$$f_{\text{GLLPH}}(t) = \frac{\beta c (ct)^{\beta-1}}{\left[1 + (\eta t)^{\beta}\right]^{c^{\beta}/\eta^{\beta}+1}}\exp(\mathbf{x}'\beta)\left[\left[1 + (\eta t)^{\alpha}\right]^{k^{\alpha}/\eta^{\alpha}}\right]^{\exp(\mathbf{x}'\beta)-1}. \tag{46}$$

The cdf of the GLL PH model is

$$F_{\text{GLLPH}}(t) = 1 - \left[\left[1 + (\eta t)^{\alpha}\right]^{-k^{\alpha}/\eta^{\alpha}}\right]^{\exp(\mathbf{x}'\beta)}. \tag{47}$$
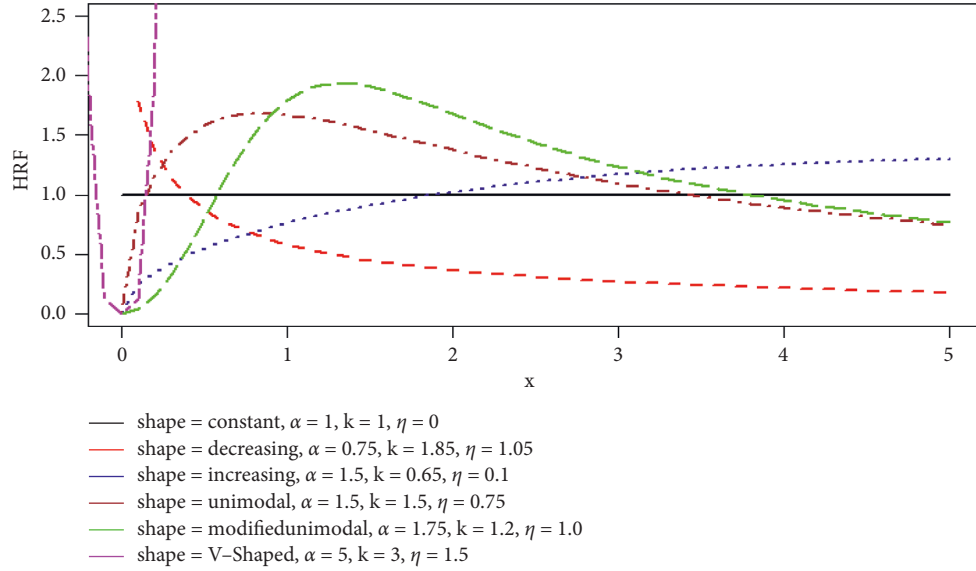
FIGURE 1: Visual representation for the different hazard rate shapes of the GLL distribution with different values of the parameters.

The chf of the GLL PH model is

$$H_{\text{GLLPH}}(t) = \exp(\mathbf{x}'\beta)\frac{k^{\alpha}}{\eta^{\alpha}} \log\left[1 + (\eta t)^{\alpha}\right]. \qquad (48)$$

6. Model Inference

We discuss the classical approach (using maximum likelihood (MLE)) and Bayesian approach (assuming noninformative priors) estimation techniques for the proposed parametric PH model parameters in this section.

6.1. MLE for Right Censored Survival Data. We examine the challenge of estimating the proposed model's distributional parameters and regression coefficients for right-censored survival data in this section. Because of its appealing qualities, such as consistency, asymptotic efficiency, asymptotic unbiasedness, and asymptotic normality, MLE is one of the most common strategies for estimating the parameters of hazard-based regression models. Let there be n individuals with lifetimes represented by $T_1, T_2, \ldots, T_n$. Assuming that the data are subject to right censoring, we observe $t_i = \min(T_i, C_i)$, where $C_i > 0$ corresponds to a potential censoring time for individual i. Allow $\delta_i = \mathrm{I}(T_i, C_i)$ that equals 1 if $T_i \leq C_i$ and 0 otherwise.

Suppose that a right-censored random sample with data $D = (t_i, \delta_i, \mathbf{x_i})$, $i = 1, 2, \ldots, n$, is available, where t_i is the censoring time or a survival time according to whether $\delta_i = 0$ or 1, respectively, and $\mathbf{x_i} = x_1, x_2, \ldots, x_n$ is an $n \times 1$ column vector of external covariates for the ith individual, ϑ is the vector of parameters associated with the baseline distribution, and β is the vector of regression coefficients. When the parametric PH model is considered, the censored likelihood function can be expressed as

$$L(\vartheta, \beta | \mathbf{D}) = \prod_{i=1}^{n} \left[f(t_i | \vartheta, \beta, \mathbf{x})\right]^{\delta_i} \left[s(t_i | \vartheta, \beta, \mathbf{x})\right]^{1-\delta_i},$$

$$= \prod_{i=1}^{n} \left[h(t_i | \vartheta, \beta, \mathbf{x}).S(t_i | \vartheta, \beta, \mathbf{x})\right]^{\delta_i} \left[s(t_i | \vartheta, \beta, \mathbf{x})\right]^{1-\delta_i},$$

$$= \prod_{i=1}^{n} \left[h(t_i | \vartheta, \beta, \mathbf{x})\right]^{\delta_i} \left[s(t_i | \vartheta t, n\beta q, hx)\right]$$

$$= \prod_{i=1}^{n} \left[h(t_i | \vartheta, \beta, \mathbf{x})\right]^{\delta_i} \exp\left[-\int_0^t h(u)du\right]$$

$$= \prod_{i=1}^{n} \left[h(t_i | \vartheta)\exp(\mathbf{x}'\beta)\right]^{\delta_i} \exp\left[-[H(t_i | \vartheta)\exp(\mathbf{x}'\beta)]\right].$$

$$(49)$$

An iterative optimization procedure (e.g., Newton–Raphson algorithm) can be used to obtain the maximum likelihood estimation $\widehat{\vartheta}$ of ϑ. Hypothesis testing and interval estimations of model parameters are possible due to the MLEs' approaching normality [7]. The natural logarithm of the likelihood function, so-called log-likelihood function can be written as follows:

$$\ell(\vartheta, \beta | \mathbf{D}) = \sum_{i=1}^{n} \delta_i \log\left[h_0(t_i | \vartheta) + \mathbf{x_i'}\beta\right] - \sum_{i=1}^{n} H_0(t_i | \vartheta)\exp(\mathbf{x_i'}\beta),$$

$$(50)$$

where β is a vector of the regression coefficients and $\vartheta' = (k, \alpha, \eta)$ is the vector of the baseline distributional parameters.

In our case, if we assume that $a = \sum_{i=1}^{n} \delta_i, p_i = \exp(\mathbf{x_i'}\beta)$ and $q_i = (\eta t_i)^{\alpha}$. Use (36) for $h_0(.)$ and note that $H_0(t; \theta) = \int_0^t h(u)du$ is the baseline cumulative hazard rate function as given by (37). The full log-likelihood function of the GLL PH model can be expressed as follows:

$$\ell(\vartheta|t) = a\log\alpha + a\alpha\log k + (\alpha - 1)\sum_{i=1}^{n}\delta_i\log t_i$$

$$- \sum_{i=1}^{n}\delta_i\log(1 + q_i) + a\log p_i - \left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}p_i\log(1 + q_i).$$

$$(51)$$

$$\frac{\partial\ell(\tau|t)}{\partial\alpha} = \frac{a}{\alpha} + a\log p_i + \sum_{i=1}^{n}\delta_i\log t_i - \frac{1}{\alpha}\sum_{i=1}^{n}\delta_i q_i\left[\frac{\log q_i}{(1 + q_i)}\right],$$

$$- \left(\frac{k}{\eta}\right)^{\alpha}\left(\frac{1}{\alpha}\right)\sum_{i=1}^{n}p_i q_i\left[\frac{\log q_i}{(1 + q_i)}\right] - \left(\frac{k}{\eta}\right)^{\alpha}\log\left(\frac{k}{\eta}\right)\sum_{i=1}^{n}p_i\log(1 + q_i),$$

$$\frac{\partial\ell(\tau|t)}{\partial\eta} = -\left(\frac{\alpha}{\eta}\right)\sum_{i=1}^{n}\delta_i\left[\frac{q_i}{1 + q_i}\right] - \left(\frac{\alpha}{\eta}\right)\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}p_i\left[\frac{q_i}{1 + q_i}\right] - \left(\frac{\alpha}{\eta}\right)\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}p_i\log\left(1 - \left[\frac{q_i}{1 + q_i}\right]\right),$$

$$(52)$$

$$\frac{\partial\ell(\tau|t)}{\partial k} = \frac{a\alpha}{k} - \left(\frac{\alpha}{k}\right)\left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}p_i\log(1 + q_i),$$

$$\frac{\partial\ell(\tau|t)}{\partial\beta_j} = \sum_{i=1}^{n}\delta_i Z_{ij} - \left(\frac{k}{\eta}\right)^{\alpha}\sum_{i=1}^{n}p_i\log(1 + q_i)Z_{ij} \text{ for } j = 1, 2, \ldots, p.$$

To maximize log-likelihood functions, many software packages are available including proven optimization algorithms.

6.2. Bayesian Inference.
In this section, Bayesian inference was used to estimate distributional parameters and regression coefficients using objective (or noninformative) priors to obtain proper posterior distributions.

6.2.1. Priors for the Model Parameters.
The specification of a prior distribution is a crucial aspect of any Bayesian inference. In parametric survival regression models, this is especially true. As a result, the prior scenario is built in this study using a noninformative independent prior for the parameters. The marginal prior distribution for every regression coefficient $\beta_m, m = 1, \ldots, 5$, is prompted as a normal distribution centred at zero and with a small precision, $N(0, 0.001)$; on the other hand, a gamma distribution, $gamma(10, 10)$, is chosen as the marginal prior distribution for the parameters of the GLL PH model due to the versatility of gamma distribution that include the noninformative priors (uniform) on the shape parameters. Many research publications in the literature, such as Danish and Aslam [23, 24], considered the assumption of the gamma priors for the baseline hazard parameters of PH models. Alvares et al. [1] took the assumption of independent gamma priors for the baseline hazard parameters of eight different parametric survival models. Muse et al. [22]

used the assumption of independent gamma priors for the baseline hazard parameters of the of the generalized log-logistic AFT model, and other researchers take these priors into account.

For the baseline parameters of the GLL-PH model, we assume independent gamma priors.

$$p(\alpha) \sim G(a_1, b_1) = \frac{b_1^{a_1}}{\Gamma(a_1)}\alpha^{a_1 - 1}e^{-b_1\alpha}; \ a_1, b_1, \alpha > 0,$$

$$p(\eta) \sim G(a_2, b_2) = \frac{b_2^{a_2}}{\Gamma(a_2)}\eta^{a_2 - 1}e^{-b_2\eta}; \ a_2, b_2, \eta > 0, \quad (53)$$

$$p(k) \sim G(a_3, b_3) = \frac{b_3^{a_3}}{\Gamma(a_3)}k^{a_3 - 1}e^{-b_3 k}; \ a_3, b_3, k > 0.$$

Prior to that, we had the regression coefficients (assuming a normal distribution).

$$p(\beta') \sim N(a_4, b_4). \quad (54)$$

The density function of the combined prior distribution of all unknown parameters and the regression coefficients are given as

$$p(\alpha, k, \eta, \beta') = p(\alpha)p(\eta)p(k)p(\beta'). \quad (55)$$

6.2.2. The Likelihood Function.
Unfortunately, the likelihood function of this generalized model is not implemented

To obtain the MLE's of $\theta' = (k, \alpha, \eta)$ and β', we can maximize (51) directly with respect to (k, α, η) and β' or we can solve the nonlinear equations below or the 1st derivative of the log-likelihood function. The 1st derivatives of the log-likelihood function are

in BUGS and JAGS syntax. To generate the likelihood function, we use the "zero's trick" method that become popular in survival analysis and relies on Poisson modelling of expanded or reconstructed data [1]. The zero's trick approach works on the assumption that perhaps the contribution of a Poisson (λ) observable of zero is $\exp(-\lambda)$; if we set $\lambda = -\log(f(t_i|\vartheta, \beta, \mathbf{x}))$ with observable data as a vector of $0's$, we receive the right contributions of the proposed model [18].

6.2.3. The Posterior Distribution.
The joint posterior density function is equal to the multiplication of the prior distribution $p(\alpha, k, \eta, \boldsymbol{\beta}')$ and the likelihood function the joint

posterior density function of the parameters $\alpha, k, \eta,$ $and\ d\beta'$ of GLL PH model given the data can be expressed using Bayes' theorem as

$$p(\alpha, k, \eta, \boldsymbol{\beta}'|x) \propto p(\alpha, k, \eta, \boldsymbol{\beta}')L(\alpha, k, \eta, \boldsymbol{\beta}'),$$
$$p(\alpha, k, \eta, \boldsymbol{\beta}'|x) \propto p(\alpha)p(\eta)p(k)p(\boldsymbol{\beta}')L(\alpha, k, \eta, \boldsymbol{\beta}'), \tag{56}$$

where the first four terms on the equation represent the prior specification for the unknown parameters and are assumed to be independent and $L(\alpha, k, \eta, \beta')$ is the likelihood function expressed as follows:

$$L(\alpha, k, \eta, \boldsymbol{\beta}') = \prod_{i=1}^{n} \left[\frac{\alpha k (kx)^{\alpha-1}}{[1+(\eta t)^{\alpha}]} \exp(\mathbf{x}'\beta) \right]^{\delta_i} \left[\exp(\mathbf{x}'\beta) \frac{k^{\alpha}}{\lambda^{\alpha}} \log[1+(\lambda x)^{\alpha}] \right],$$

$$p(\alpha, k, \eta, \boldsymbol{\beta}'|x) \propto \left\{ \prod_{j=0}^{\mathbf{p}} \boldsymbol{\pi}(\boldsymbol{\beta_j}) \right\} \alpha^{a_1+n-1} \eta^{a_2+n-1} k^{a_3+n-1} e^{-(b_1\alpha+b_2\eta+b_3 k)} L(\alpha, k, \eta, \boldsymbol{\beta}'). \tag{57}$$

The marginal distributions of the model parameters and the normalising joint posterior density function are difficult to calculate analytically, requiring high-dimensional integration and no close form inferences. To obtain estimates, we use McMC simulation methods, which involve sampling from the posterior distribution through using the Metropolis–Hastings Algorithm.

7. Simulation Study

In this section, we undertake an extensive simulation investigation to demonstrate the proposed parametric proportional hazard model's good Bayesian features. The parameter values are chosen to construct situations that mimic cancer population studies using a cancer that is severe (with a lower five-year survival rate), such as lung cancer [9, 25]. We demonstrate parameter estimation, the effect of censoring proportions, and sample sizes on inference in more detail.

7.1. Generating Survival Data from the PH Model.
To simulate survival data for the GLL PH model, we use the inversion technique [40, 41] to generate survival data. This strategy is based on the link between a survival random variable's cumulative hazard rate function and a standard uniform random variable. When the cumulative hazard rate function has a closed form expression, it may be immediately applied, inverted, and readily implemented with R [26]. The censoring rates were estimated using administrative censorship at (1) Tc = 5 years, which resulted in around 20% censoring in all sets, and (2) Tc = 3 years, which resulted in about 30% censoring in all sets.

For the purposes of this simulation, we assume that survival times are distributed using the generalized log-

logistic distribution (α, η, k). Using the reverse chf given in equation (41), lifetime data can be simulated as follows:

$$T = H_0^{-1} \left\{ \frac{\left(e^{(\eta/k)^{\alpha}\left[-\log(1-U)/e^{\beta x_i} \right]} - 1 \right)^{1/\alpha}}{\eta} \right\}, \tag{58}$$

where a, η, and $k > 0$.

7.2. Simulation Design.
The simulation analysis was carried out by conducting a series of simulations with different sample sizes ($n = 100$, and 300) sets and censoring proportions (Tc = 20 and 30 percentages), all based on the PH model in equation (1). The GLL PH model's true parameter vector is set as follows: (1) set I: distributional parameter values ($\alpha = 1.5$, $k = 0.75$, and $\eta = 1.25$) and covariates $\beta = (0.75, -0.75, 0.5)$, (2) set II: distributional parameter values ($\alpha = 1.5$, $k = 0.95$, and $\eta = 1.5$) and covariates $\beta = (0.75, -0.75, 0.5)$.

The values of the covariates were simulated as follows: (1) combination of uniform distributions with 0.25 probability on (30, 65), 0.35 probability on (65, 75), and 0.40 probability on (75, 85) years old was used to simulate the continuous covariate "age," and (2) the binary covariates "treatment" and "gender" were both simulated using a 0.5 binomial distribution. We recommend that the reader can refer [9] for further details.

7.3. Posterior Analysis of the Simulated Data.
We fitted the proposed PH model with GLL baseline hazard to assess its Bayesian properties in the simulation sets. With all censoring rates and different sample sizes, each simulation set was used to estimate the proposed PH model. JAGS software

TABLE 1: Simulation results from a GLL PH framework with distributional parameters ($\alpha = 1.5$, $k = 0.75$, and $\eta = 1.25$), covariates $\beta = (0.75, -0.75, 0.5)$, and $n = 100$.

True value ($\widehat{\theta}$)	Posterior mean ($\widehat{\theta}$)	Bias	Naïve SE	MSE	CP	$\widehat{R}$	No. of eff
			Posterior properties				
			$C = 20\%$				
$\alpha = 1.50$	1.506	0.006	0.001	0.032	0.032	1.000	3782
$\beta_1 = 0.75$	0.837	0.087	0.002	0.057	0.935	1.001	3740
$\beta_2 = -0.75$	−0.695	0.055	0.001	0.011	0.945	1.000	3720
$\beta_3 = 0.50$	0.480	0.020	0.002	0.049	0.920	1.002	3700
$\eta = 1.25$	1.431	0.181	0.002	0.107	0.890	1.000	4039
$k = 0.75$	0.720	0.030	0.001	0.013	0.935	1.001	4039
			$C = 30\%$				
$\alpha = 1.50$	1.463	−0.037	0.001	0.029	0.935	1.000	3802
$\beta_1 = 0.75$	0.872	0.122	0.002	0.072	0.880	1.000	3823
$\beta_2 = -0.75$	−0.727	0.023	0.001	0.008	0.945	1.001	3761
$\beta_3 = 0.50$	0.501	0.001	0.002	0.060	0.997	1.000	3700
$\eta = 1.25$	1.575	0.325	0.002	0.193	0.851	1.002	3865
$k = 0.75$	0.567	−0.183	0.001	0.045	0.911	1.000	4084

[27] was used to approximate posterior distributions using three parallel chains with 40,000 iterations each plus another 3,000 for the burn-in period. To minimize autocorrelation in the sequences, the chains were thinned further by storing every 10th draw.

7.4. Measures of Performance.

The actual mean, standard deviation (SD), Naive standard error, bias, percentage of bias, coverage probability (CP), potential scale reduction factor ($\widehat{R}$), and the effective number of separate simulation draws were used to test the posterior distribution stability for the suggested PH model.

7.4.1. Evaluating the Performance of the Estimators.

We calculate the bias of the estimators using:

$$\text{Bias} (\widehat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} (\widehat{\theta} - \theta). \tag{59}$$

An underestimation is indicated by a negative bias, whereas an overestimation is shown by a positive bias.

7.4.2. Accuracy of the Estimators.

The mean square error (MSE) is a good indicator of overall accuracy and is calculated as follows:

$$\text{MSE} (\widehat{\theta}) = \frac{1}{N} \sum_{i=1}^{N} (\widehat{\theta} - \theta)^2. \tag{60}$$

This metric determines how accurate the estimates are as follows. The lower the MSE, the more accurate the estimations of impacts.

The Naive standard error, which is calculated by dividing the posterior standard deviation by the square root of the sample size, is another accuracy metric. As a result, the smaller the standard error, the larger the sample size. The Naïve SE incorporates simulation error rather than posterior uncertainty.

$$\text{Naive SE} = \frac{\text{posterior SD}}{\sqrt{n}}. \tag{61}$$

7.4.3. Coverage.

The 95 percent coverage probability (CP) is the percentage of N simulated data sets in which the true estimates were included in the 95 percent confidence interval. The more precise the estimations are, the closer the outcome is to a 95 percent coverage probability. The following is how CP is expressed:

$$CP = \widehat{\theta} \mp 1.96 \times SE (\widehat{\theta}). \tag{62}$$

7.4.4. Convergence Diagnostics.

Quantitatively, Gelman et al. [28] recommended that the acceptable limit of multivariate potential scale reduction factor (MPSRF) and potential scale reduction factor (PSRF) be near 1 $\widehat{R} < 1.1$, and the effective number of sample size simulation draws be greater than or equal to 100 for checking the convergence of McMC simulations. It is clear from the summary characteristics (Tables 1–4) that the PSRF is less than 1.1, that number of sample size simulation draws is larger than 100, and that Naive SE is smaller than the standard deviations (SD) for all of the distributional parameters and regression coefficients, as expected, indicating that the McMC algorithm has converged to the posterior distribution. Trace plots, autocorrelation plots, and Gelman plot diagnostics are the most common ways to judge the convergence of a McMC simulation graphically [28]. The McMC simulation has been achieved as evidenced by the trace plot, density plot, autocorrelation plot, and Gelman diagnostic plots for each distributional parameter and regression coefficients. That is, the McMC simulation for the GLL PH model explores the target posterior distribution appropriately.

7.5. Simulation Results.

Tables 1–4 shows the simulation results for the posterior mean, bias, Naive standard error

TABLE 2: Simulation results from a GLL PH framework with distributional parameters ($\alpha = 1.5$, $k = 0.75$, and $\eta = 1.25$), covariates $\beta = (0.75, -0.75, 0.5)$, and $n = 300$.

True ($\hat{\theta}$) Value	Posterior mean ($\hat{\theta}$)	Bias	Naïve SE	MSE	CP	$\hat{R}$	No. of eff
			Posterior properties				
			C = 20%				
$\alpha = 1.50$	1.449	−0.001	0.001	0.017	0.991	1.000	4017
$\beta_1 = 0.75$	0.712	−0.038	0.001	0.019	0.946	1.000	3700
$\beta_2 = -0.75$	−0.723	0.027	0.000	0.003	0.956	1.000	3761
$\beta_3 = 0.50$	0.483	−0.017	0.001	0.016	0.962	1.000	3782
$\eta = 1.25$	1.309	0.059	0.002	0.070	0.923	1.001	4609
$k = 0.75$	0.731	−0.019	0.001	0.012	0.941	1.001	4609
			C = 30%				
$\alpha = 1.50$	1.527	0.027	0.001	0.021	0.945	1.000	4174
$\beta_1 = 0.75$	0.726	−0.024	0.001	0.023	0.954	1.000	3660
$\beta_2 = -0.75$	−0.752	0.002	0.000	0.004	0.975	1.001	3802
$\beta_3 = 0.50$	0.445	−0.055	0.001	0.023	0.937	1.001	3802
$\eta = 1.25$	1.437	0.187	0.002	0.123	0.911	1.002	4434
$k = 0.75$	0.847	0.097	0.001	0.023	0.907	1.003	4792

TABLE 3: Simulation results from a GLL PH framework with distributional parameters ($\alpha = 1.75$, $k = 0.95$, and $\eta = 1.5$), covariates $\beta = (0.5, -0.85, 0.5)$, and $n = 100$.

True ($\hat{\theta}$) Value	Posterior mean ($\hat{\theta}$)	Bias	Naïve SE	MSE	CP	$\hat{R}$	No. of eff
			Posterior properties				
			C = 20%				
$\alpha = 1.75$	1.718	−0.032	0.002	0.038	0.942	1.000	3865
$\beta_1 = 0.50$	0.523	0.023	0.002	0.051	0.955	1.000	3823
$\beta_2 = -0.85$	−0.817	−0.033	0.001	0.010	0.946	1.000	3720
$\beta_3 = 0.50$	0.489	−0.011	0.002	0.050	0.981	1.000	3740
$\eta = 1.50$	1.441	−0.059	0.002	0.068	0.931	1.000	4084
$k = 0.95$	0.828	−0.122	0.001	0.147	0.925	1.001	4084
			C = 30%				
$\alpha = 1.75$	1.717	−0.033	0.002	0.044	0.939	1.000	3802
$\beta_1 = 0.50$	0.577	0.077	0.002	0.063	0.943	1.000	3823
$\beta_2 = -0.85$	−0.833	−0.017	0.001	0.009	0.971	1.000	3761
$\beta_3 = 0.50$	0.474	−0.026	0.002	0.058	0.952	1.000	3700
$\eta = 1.50$	1.625	0.125	0.002	0.143	0.919	1.002	3865
$k = 0.95$	0.778	−0.172	0.001	0.213	0.908	1.001	4084

TABLE 4: Simulation results from a GLL PH framework with distributional parameters ($\alpha = 1.75$, $k = 0.95$, and $\eta = 1.5$), covariates $\beta = (0.5, -0.85, 0.5)$, and $n = 300$.

True ($\hat{\theta}$) value	Posterior mean ($\hat{\theta}$)	Bias	Naïve SE	MSE	CP	$\hat{R}$	No. of eff
			Posterior properties				
			C = 20%				
$\alpha = 1.75$	1.756	0.006	0.001	0.023	0.978	1.000	3951
$\beta_1 = 0.50$	0.503	0.003	0.001	0.040	0.991	1.000	3761
$\beta_2 = -0.85$	−0.827	−0.023	0.000	0.003	0.963	1.000	3761
$\beta_3 = 0.50$	0.505	0.005	0.000	0.045	0.987	1.000	3740
$\eta = 1.50$	1.519	0.019	0.002	0.107	0.942	1.000	4458
$k = 0.95$	0.973	0.023	0.001	0.013	0.941	1.001	4458
			C = 30%				
$\alpha = 1.75$	1.811	0.061	0.001	0.091	0.935	1.000	4011
$\beta_1 = 0.50$	0.612	0.112	0.001	0.129	0.880	1.000	3978
$\beta_2 = -0.85$	−0.815	−0.035	0.000	0.004	0.945	1.000	4011
$\beta_3 = 0.50$	0.521	0.021	0.001	0.063	0.997	1.000	3789
$\eta = 1.50$	1.531	0.031	0.002	0.171	0.851	1.001	4458
$k = 0.95$	0.990	0.040	0.002	0.145	0.911	1.002	4565
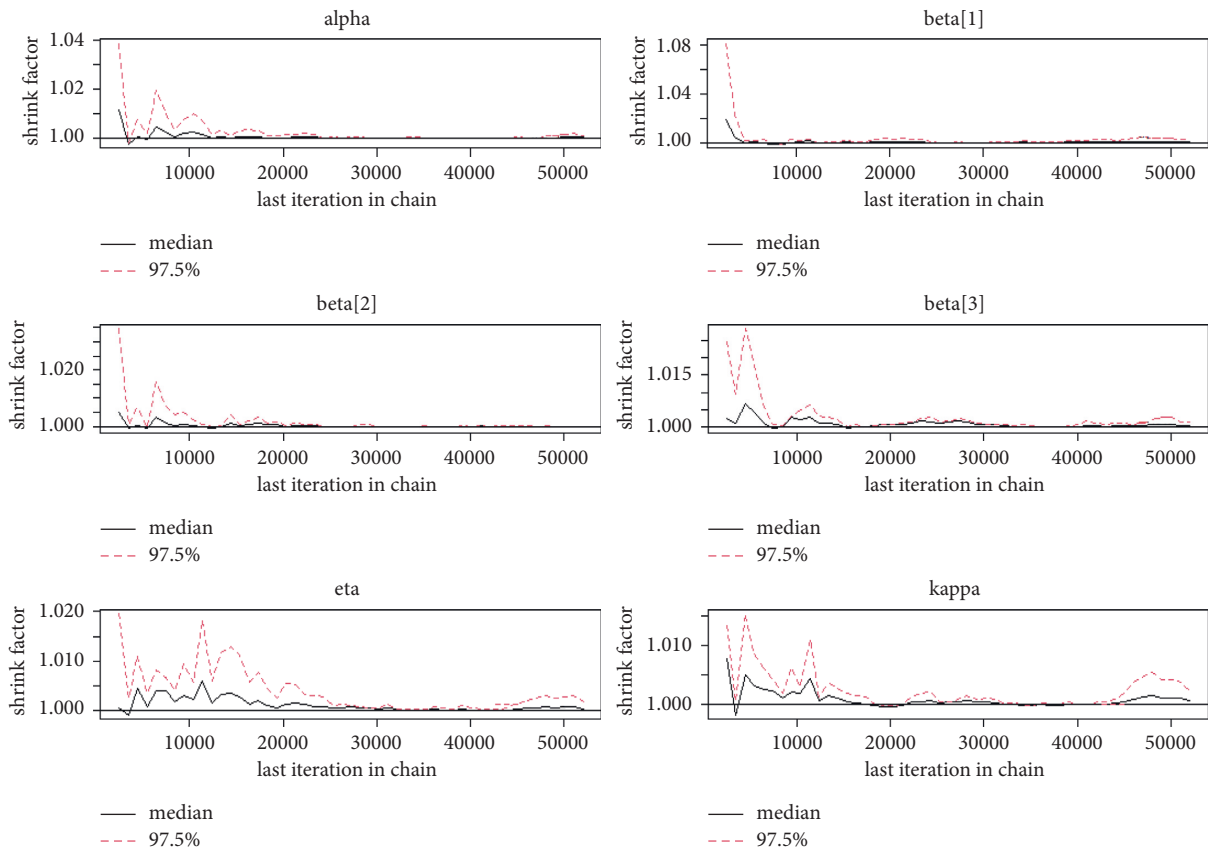
Figure 2: Gelman diagnostics from a GLL PH framework with distributional parameters ($\alpha = 1.5$, $k = 0.75$, and $\eta = 1.25$), covariates $\beta = (0.75, -0.75, 0.5)$, and $n = 300$ and censoring proportion for 20 percentage.
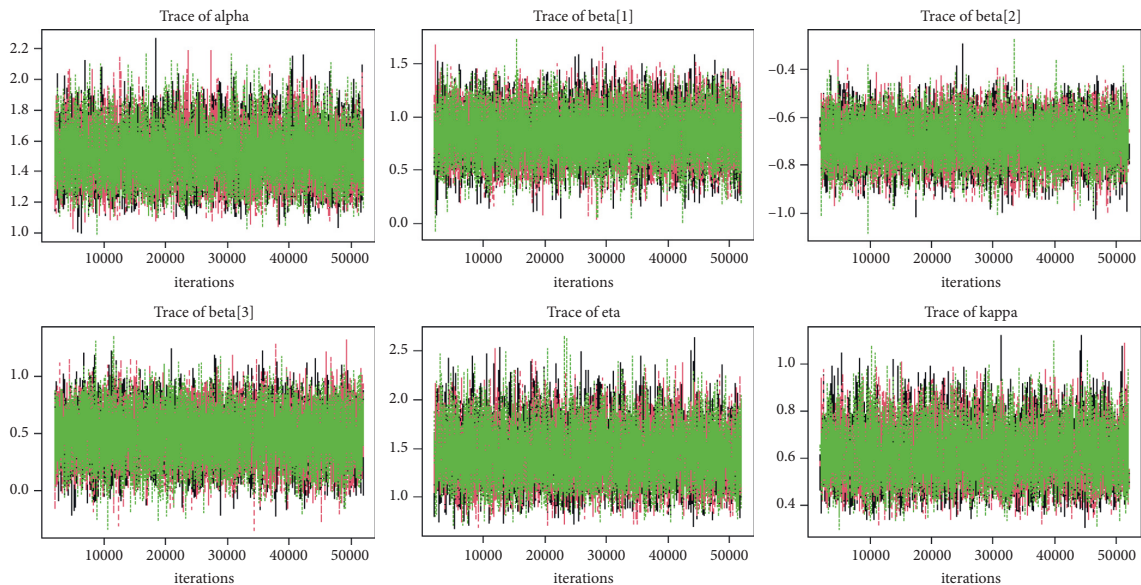
Figure 3: Trace plots from a GLL PH framework with distributional parameters ($\alpha = 1.5$, $k = 0.75$, and $\eta = 1.25$), covariates $\beta = (0.75, -0.75, 0.5)$, and $n = 300$ and censoring proportion for 20 percentage.

(SE), mean square error (MSE), coverage probability (CP), Gelman–Rubin diagnostic ($\widehat{R}$), and the number of sample size simulation draws (no. of Eff) of the proposed PH model, and Figures 2–5 shows the visual summary for the convergence diagnostics.

Based on these findings, we may deduce that, as the sample size grows, the biases and MSE of the estimators decrease; also, the censoring proportion impacts the bias and MSE of the estimators, with larger censoring rates increasing the bias and MSE. The Gelman–Rubin diagnostic, on the
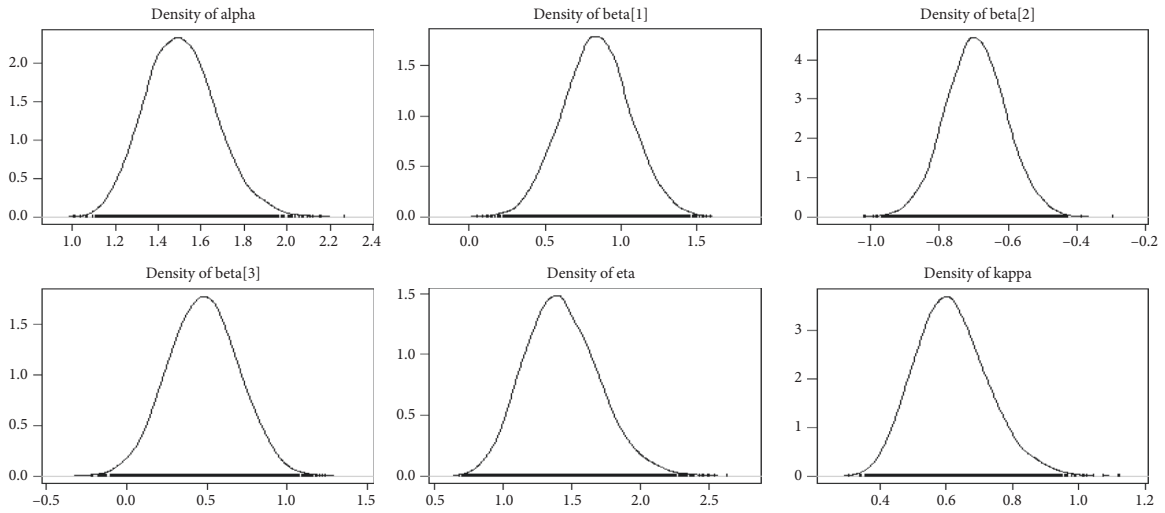
Figure 4: Kernel density plots from a GLL PH framework with distributional parameters ($\alpha = 1.5$, $k = 0.75$, and $\eta = 1.25$) and covariates $\beta = (0.75, -0.75, 0.5)$, and $n = 300$ and censoring proportion for 20 percentage.
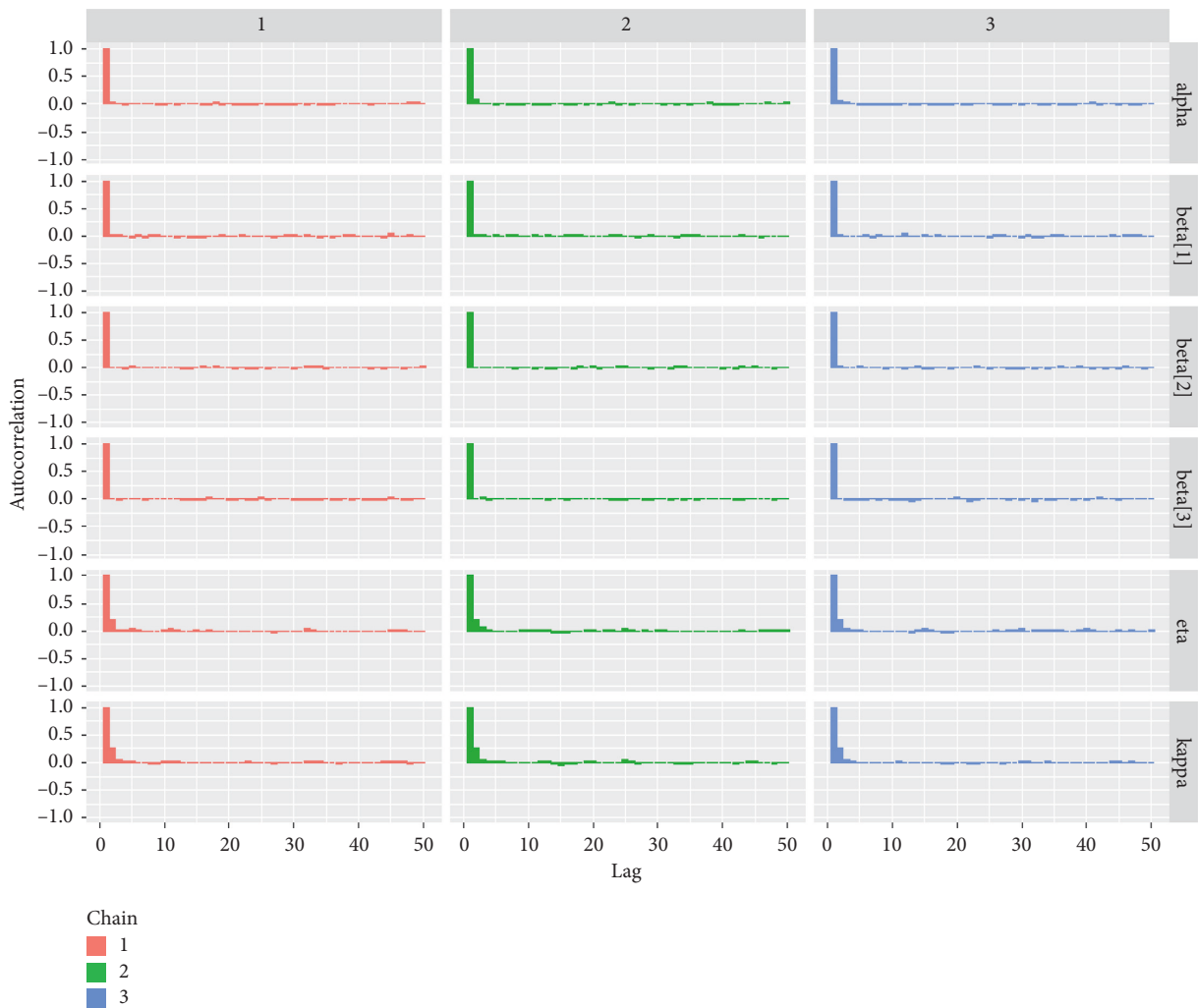
Figure 5: Autocorrelation plots from a GLL PH framework with distributional parameters ($\alpha = 1.5$, $k = 0.75$, and $\eta = 1.25$), covariates $\beta = (0.75, -0.75, 0.5)$, and $n = 300$ and censoring proportion for 20 percentage.

other hand, as well as the number of efficiency sample size draws show that convergence has been attained. However, the estimators' coverage probability was close to 95%.

8. Practical Illustrations

In this section, two real-life survival data sets dealing with right-censored cancer data sets were considered to demonstrate the flexibility and applicability of the proposed GLL PH in modelling different survival data sets with different hazard rate shapes.

8.1. Data Set I: Lung Cancer Survival Data

8.1.1. Data Description. In this section, we reanalyse the data set reported in [29] which is available in the R package survival. The Veterans Administration Lung Cancer Study Group followed up on $n = 137$ patients in this data set. For this clinical investigation, the censorship rate is around 6.5 percent (9 observations out of 137 were censored). The response and exploratory factors in this clinical trial are the time until death (in days), the number of months from diagnosis to study enrolment (diagt), age (in years), a history of previous lung cancer therapy (prior), and the trt = (treatment = conventional chemotherapy).

8.1.2. Hazard Rate Shape. The hazard rate function appears to be unimodal or decreasing in Figure 6 based on the TTT plot (careful inspection reveals a slight indication of unimodality). The data could be evaluated with a model like the log-logistic distribution, which can accommodate decreasing or unimodal hazard rate forms. However, because the classical LL distribution is not closed under the PH framework, we employ the GLL distribution, which is closed and can encompass various hazard rate shapes. The box plot, histogram, and TTT plots are shown in Figure 6.

8.1.3. Proportionality Assumption. There are two widely used methods for assessing the PH assumption: (1) graphical diagnostics based on (a) time-dependent variables [7] and (b) standardized Schoenfeld residuals [30] and (2) statistical tests. The standardized Schoenfeld residuals are used in this section to evaluate the PH assumption of the Cox model for each covariate included in the model. Based on Figure 7 and the significance threshold of 5%, there is no evidence to reject the proportional hazards assumption. As a result, we anticipate that the GLL PH model will provide a good fit when compared to the other existing parametric PH model employed in this study.

8.1.4. Posterior Analysis. In this paper, we assume the noninformative independent framework with a normal prior $N(0, 0.001)$ for $\beta ıs$ (regression coefficients) and an independent gamma prior for the distributional parameters $\alpha \sim G(a_1, b_1)$, $\eta \sim G(a_2, b_2)$, and $k \sim G(a_3, b_3)$ with hyperparameter values $(a_1 = b_1 = a_2 = b_2 = a_3 = b_3 = 10)$.

(1) Numerical Summary. We looked at various quantities of interest and their numerical values using the McMC sample of posterior properties for the generalized log-logistic proportional hazard model using the lung cancer data in this section.

The posterior summaries for the generalized log-logistic PH model parameters using Veterans lung cancer data sets are illustrated in Table 5. The probability that the corresponding parameter is +ve is given in the last row of Table 5. A posterior probability of 0.5 indicates that a positive parameter value is as likely as a negative one. Once we've saved the posterior sample for each model parameter, we can compute the posterior probability, for example, for β_1, using mean $(\beta_1 > 0)$.

(2) Visual Summary. We looked at density strip plots, trace plots, Gelman–Rubin diagnostic plots, Ergodic mean plots, and autocorrelation diagnostic plots in this section to get a visual description of the posterior properties. These plots and graphs provide a nearly comprehensive representation of the parameters' posterior uncertainty for the application of the lung cancer data sets.

(3) Density Plots. Density can be compared to the fundamental shapes associated with typical analytic distributions, and density plots can reveal behaviour in the tails, skewness, existence of multimodal behaviour, and data outliers. Figure 8 illustrates the density plots for the GLL PH model parameters.

(4) Time-Series Plots. One of the most common diagnostics of an McMC simulation is a time series plot (or trace plot) [28]. Figure 9 shows that the McMC sampling process converges to the joint posterior distribution with no periodicity. As a result, we can say that the chains have converged.

(5) Brooks–Gelman–Rubin (BGR) Convergence Diagnostic. Gelman and Rubin [31] propose a convergence diagnostic technique to check the McMC algorithms simulation and is based on within chain variance and between chain variance. Gelman et al. [28] suggested that the limit of acceptance of potential scale reduction factor (PSRF) to be less than 1.1. Figure 10 shows us that both PSRF and MPSRF are less than 1.1.

(6) Running Mean Plots. The running mean (also referred to Ergodic mean) is a well-known convergence diagnostic for McMC algorithms. The Ergodic mean is defined as the mean of all simulated sample values of up to a specific iteration [32]. Ergodic mean is used to observe the convergence pattern of the McMC chains. Figure 11 shows us the Ergodic mean plots for the regression coefficients and the baseline hazard parameters. It is quite clear from the running mean time-series plots that the chains converge after N iterations to their mean values. However, these plots display only at the mean of the baseline hazard parameters and the regression coefficients and therefore are inadequate.

(7) Autocorrelation Plots. Although the autocorrelation plot is not strictly a convergence diagnostic tool, it does aid in indirectly assessing the convergence of the McMC simulation process [33]. Figure 12 shows the autocorrelation plots for all parameters and regression coefficients.
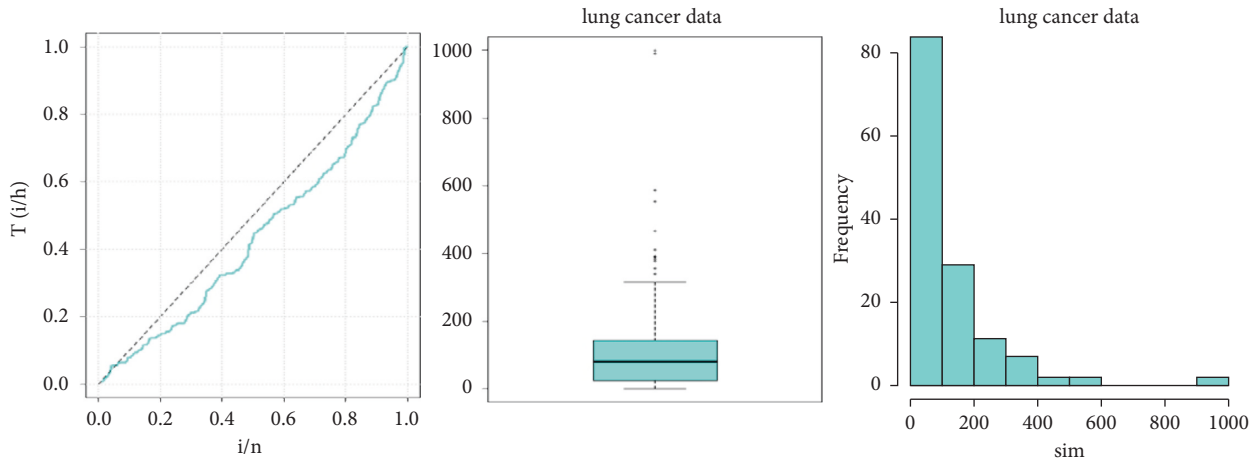
FIGURE 6: TTT plot, box plot, and the histogram for the survival times of the lung cancer data sets.

FIGURE 7: The standardized Schoenfeld residuals from the data I—lung cancer data set, taking the test p value for each covariate into account.

TABLE 5: Numerical summaries of posterior characteristics based on McMC sample of the GLL PH model for the lung cancer data set.

Characteristics	Pars						
	Alpha	β_1 (diagt)	β_2 (age)	β_3 (prior)	β_4 (trt)	Eta	Kappa
Mean	1.317	0.002	−0.024	−0.015	−0.151	0.042	0.103
SD	0.173	0.010	0.008	0.021	0.178	0.015	0.049
Naïve SE	0.001	0.0001	0.0001	0.0002	0.001	0.0001	0.0004
Time series SE	0.003	0.0001	0.0001	0.0001	0.002	0.0002	0.001
Minimum	0.813	−0.046	−0.054	−0.109	−0.890	0.007	0.019
2.5th percentile	1.023	−0.020	−0.040	−0.057	−0.500	0.019	0.040
Q1	1.194	−0.005	−0.029	−0.029	−0.271	0.031	0.068
Medium (Q2)	1.302	0.003	−0.024	−0.015	−0.150	0.040	0.092
Q3	1.422	0.010	−0.018	−0.0003	−0.029	0.051	0.125
97.5th percentile	1.697	0.021	−0.007	0.027	0.193	0.078	0.231
Maximum	2.324	0.032	0.006	0.082	0.511	0.143	0.658
Mode	1.250	0.003	−0.028	−0.015	−0.150	0.035	0.075
Variance	0.030	0.0001	0.0001	0.001	0.032	0.0002	0.002
Skewness	0.550	−0.361	0.082	−0.058	−0.027	0.957	1.656
Kurtosis	0.558	0.152	0.011	0.001	−0.009	1.510	4.992
95% credible interval	(1.023, 1.697)	−0.020, 0.021)	(−0.040, −0.007)	(−0.057, 0.027)	(−0.500, 0.193)	(0.019, 0.078)	(0.040, 0.231)
P (.>0|data)	1.000	0.598	0.003	0.244	0.199	1.000	1.000
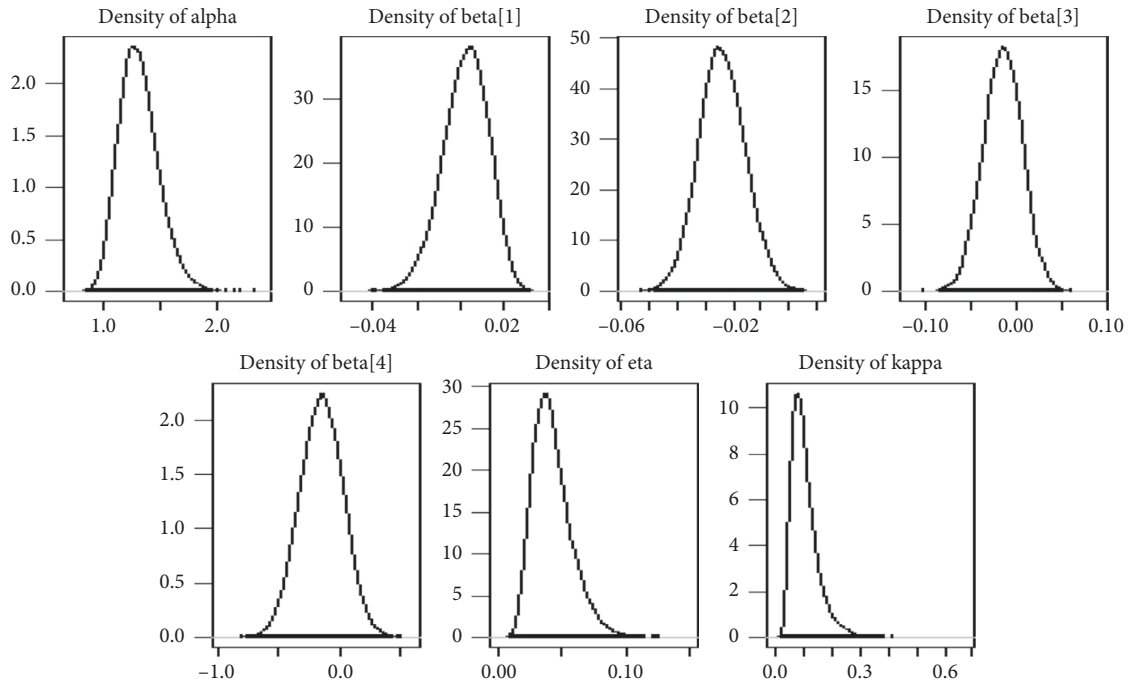
Figure 8: Density plots for regression coefficients and distributional parameters from the Veterans lung cancer data set.
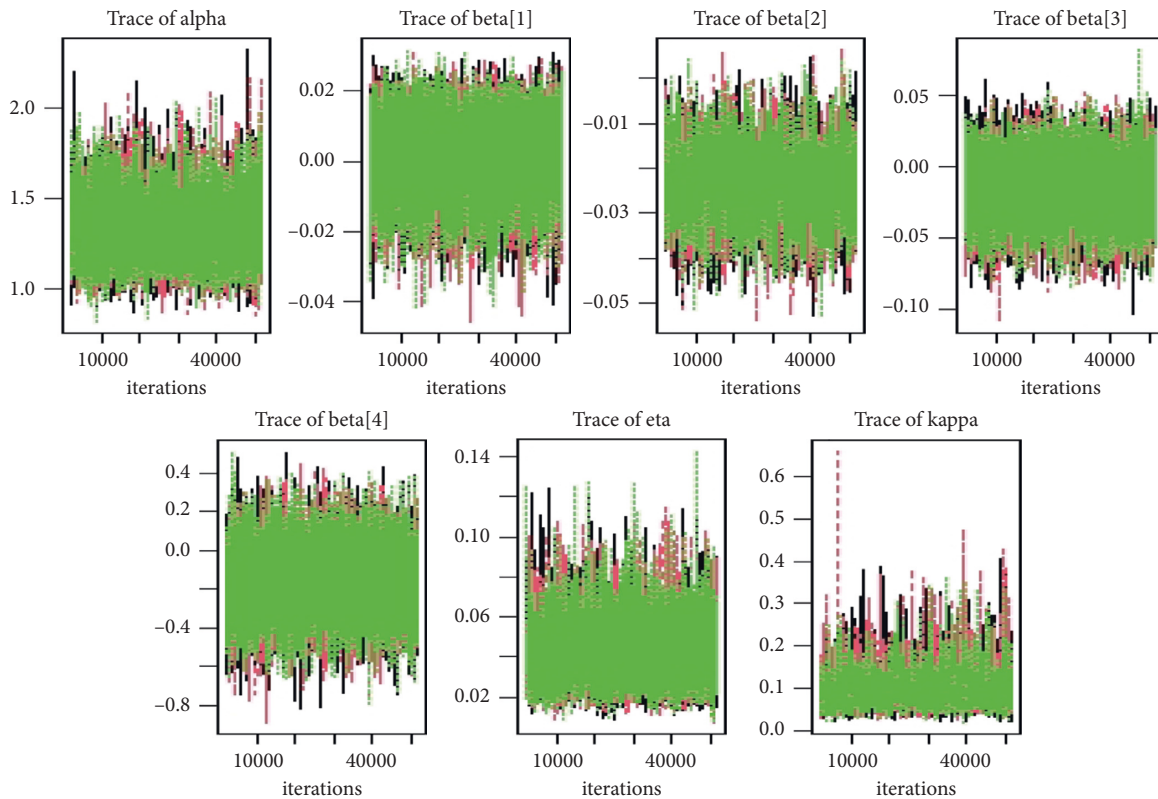
Figure 9: The time series plots for the baseline hazard parameters and the regression coefficients for the Veterans lung cancer data.

8.1.5. Convergence of McMC Algorithm for the Veterans Lung Cancer Data Set. Computational developments in the previous few decades have recently emerged as a very useful instrument for employing McMC approaches [34] and

fitting Bayesian survival regression models in time-to-event analysis. The complicated posterior distribution is sampled using the McMC algorithm. As a result, when an algorithm converges to the target posterior distribution, the Markov
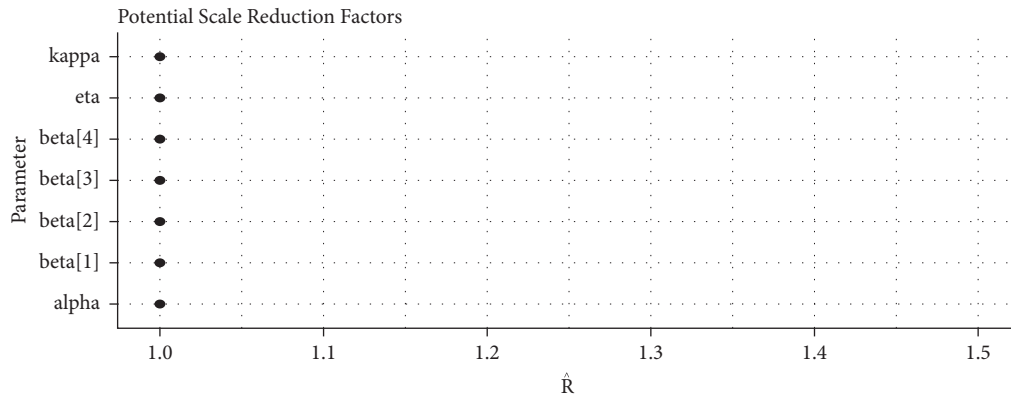
Figure 10: PSRF of the baseline hazard parameters and the regression coefficients for the Veterans lung cancer data.
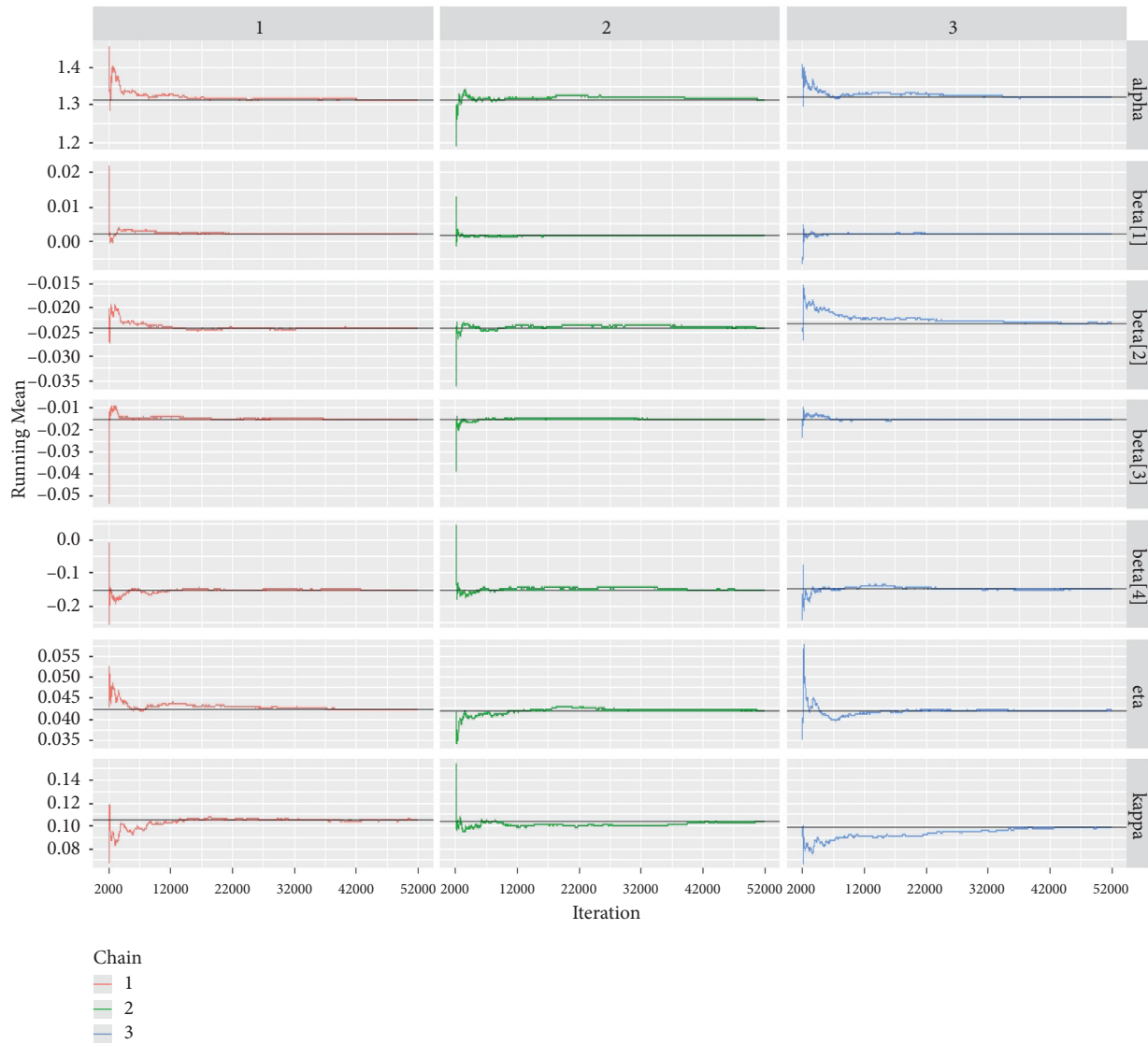
Figure 11: The running mean plots for the baseline distributional parameters and regression coefficients for the Veterans lung cancer data set.
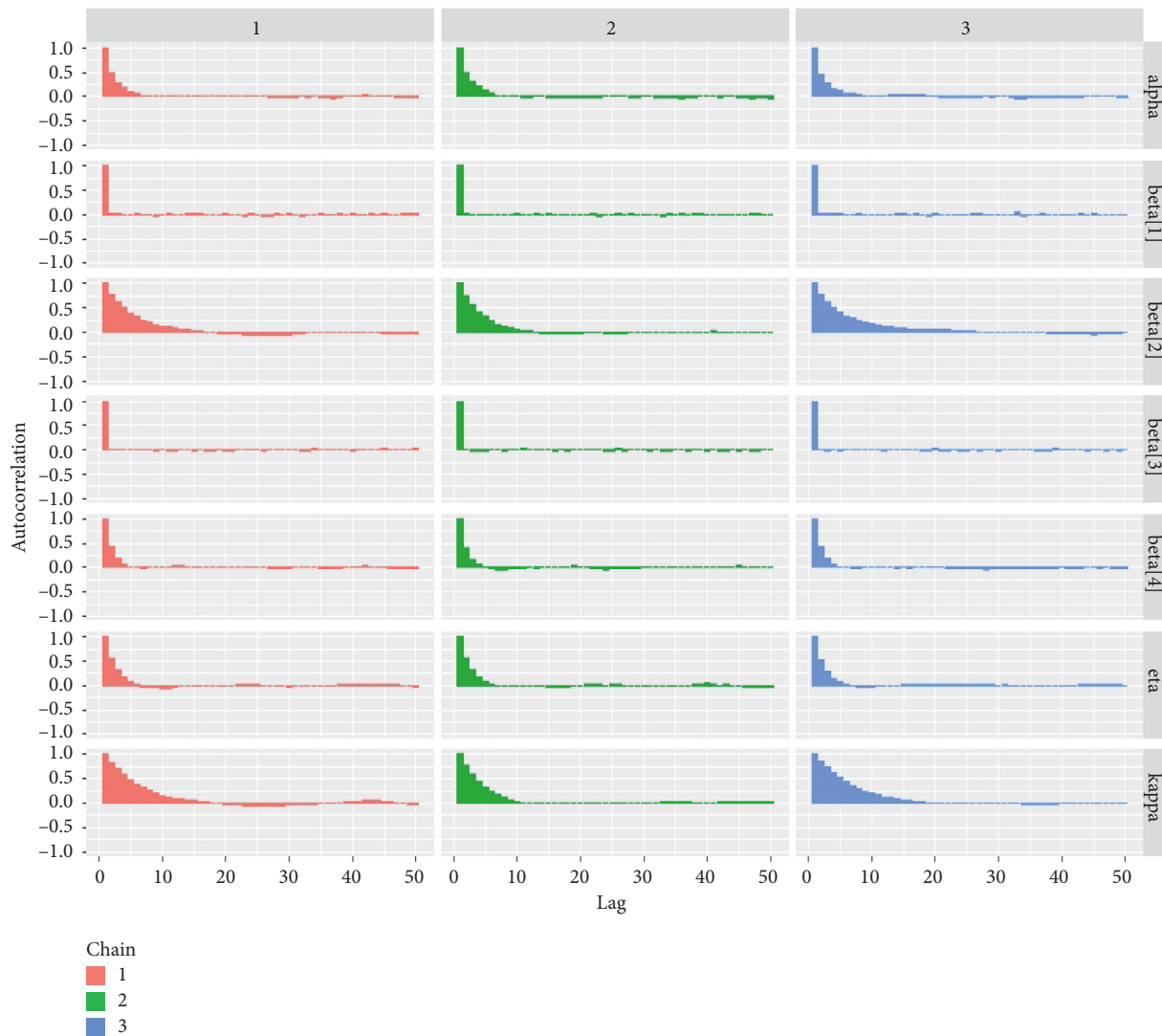
FIGURE 12: Autocorrelation plots for all the baseline distributional parameters and regression coefficients for the Veterans lung cancer data set.

chain is stationary, and adding more samples will not change the shape and position of the posterior distribution's density in a meaningful way and hence will not change the estimations or other relevant outcomes.

(1) Common Statistical Tests for Convergence Diagnostics. The convergence of the McMC algorithm was checked quantitatively using conventional statistical tests for convergence diagnostics: (1) Brooks–Gelman–Rubin diagnostics [28]; (2) Raftery and Lewi diagnostics [35]; (3) Heidelberger and Welch's diagnostic tests [36]; and (4) Geweke diagnostics [37]. For more information about these tests, we can refer to [34]. Table 6 indicates the Geweke, Raftery–Lewis, and Heidelberger–Welch diagnostics for the GLL PH model parameters.

(2) Graphical Techniques for Convergence Diagnostics. Convergence diagnostics of an McMC algorithm can be examined graphically, including: (1) time series plot; (2)

autocorrelation plot; (3) running mean plot; and (4) Gelman–Rubin plots. See Figures 9–12.

8.2. Data Set II: Larynx Cancer Data Sets

8.2.1. Data Description. Lifetimes for 90 patients with larynx-cancer, according to the stage of cancer tumour (stages I–IV) are given in Table 7. The study time or time to death are recorded in months (where, * shows us the censored time). Alvares et al. [1]; Wang et al. [8]; and Christensen et al. [19] discussed the data from different aspects under different hazard-based regression models, and the data were first reported by [38]. The survival times (in months) of patients is illustrated in Table 7.

The other covariates of the data are as follows: (1) age (in years) at diagnosis and (2) the year of diagnosis. One goal of this study was to see if the age, year of diagnosis, and stage of cancer were associated with the death of patients with laryngeal cancer.

TABLE 6: Summaries for Raftery–Lewis's diagnostic, Geweke diagnostic, and Heidelberger–Welch diagnostics test of the GLL PH model parameters for the Veterans lung cancer data set.

Parameter	Geweke diagnostic Pr > |z|	Diagnostics for the Raftery Lewis Dependency factor (I)	Diagnostics for the Heidelberger–Welch		
			Stationarity test	p value	Halfwidth test
Alpha	−0.383	2.430	Passed	0.648	Passed
β_1 (diagt)	0.820	1.030	Passed	0.337	Passed
β_2 (age)	−0.272	3.640	Passed	0.613	Passed
β_3 (prior)	−0.680	0.988	Passed	0.885	Passed
β_4 (trt)	0.608	2.120	Passed	0.112	Passed
Eta	−1.436	1.160	Passed	0.178	Passed
Kappa	−0.142	3.500	Passed	0.506	Passed

TABLE 7: Survival times (in months) of patients with larynx cancer according to stages of tumour (1–4).

Stages	Survival time (* = indicating censoring)
Stage I (33 patients)	0.6, 1.3, 2.4, 2.5*, 3.2, 3.3*, 3.5, 3.5, 4.0, 4.0, 4.3, 4.5*, 4.5*, 5.3, 5.5*, 5.9*, 5.9*, 6.0, 6.1*, 6.2*, 6.4, 6.5, 6.5*, 6.7*, 7.0*, 7.4, 7.4*, 8.1*, 8.1*, 9.6*, 10.7*
Stage II (17 patients)	0.2, 1.8, 2.0, 2.2*, 2.6*, 3.3*, 3.6, 4.0*, 4.3, 4.3*, 5.0*, 6.2, 7.0, 7.5*, 7.6*, 9.3*
Stage III (patients)	0.3, 0.3, 0.5, 0.7, 0.8, 1.0, 1.3, 1.6, 1.8, 1.9, 1.9, 3.2, 3.5, 3.7*, 4.5*, 4.8*, 4.8*, 5.0, 5.0*, 5.1*, 6.3, 6.4, 6.5*, 7.8, 8.0*, 9.3*, 10.1*
Stage IV (13 patients)	0.1, 0.3, 0.4, 0.8, 0.8, 1.0, 1.5, 2.0, 2.3, 2.9*, 3.6, 3.8, 4.3*
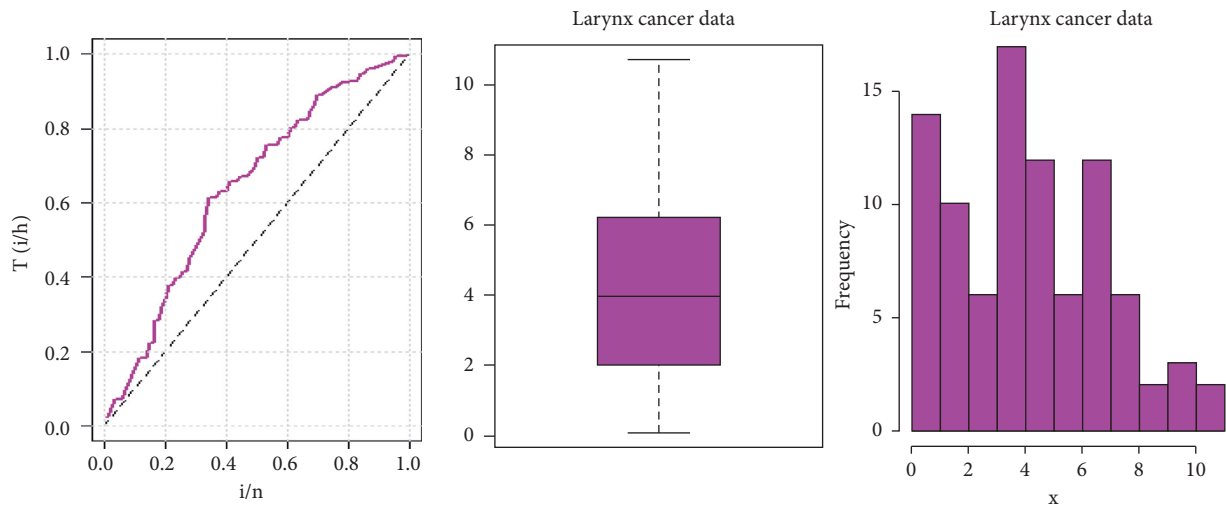
FIGURE 13: TTT plot, box plot, and the histogram for the survival times of the larynx cancer data set.

8.2.2. Hazard Rate Shape.

Based on the TTT plot, the hazard rate function is an increasing hazard in Figure 13. The data could be analyzed using a model such as the Weibull distribution, which can handle monotone hazard rate forms. We adopt the GLL distribution, which would be represented by the PH framework and can accommodate a variety of hazard rate shapes to see its applicability of the monotone (increasing) hazard rates. Figure 13 shows the box plot, histogram, and TTT plots.

8.2.3. Proportionality Assumption.

We investigated if the proportional hazards model could be used with this data set. The underlying assumption of the Cox model for each explanatory variable utilized in the model is depicted in Figure 14. With a significance level of 5%, there is no evidence to reject the PH assumption. As a result, we anticipate that the parametric PH model will provide a strong fit.

8.2.4. Posterior Analysis.

In this paper, we assume the noninformative independent framework with $N(0, 0.001)$ for $\beta's$ (regression coefficients) and an independent gamma prior for the distributional parameters $\alpha \sim G(a_1, b_1)$, $\eta \sim G(a_2, b_2)$, and $k \sim G(a_3, b_3)$ with hyperparameter values $(a_1 = b_1 = a_2 = b_2 = a_3 = b_3 = 10)$.

(1) Numerical Summary. We looked at various quantities of importance as well as their numerical values using the
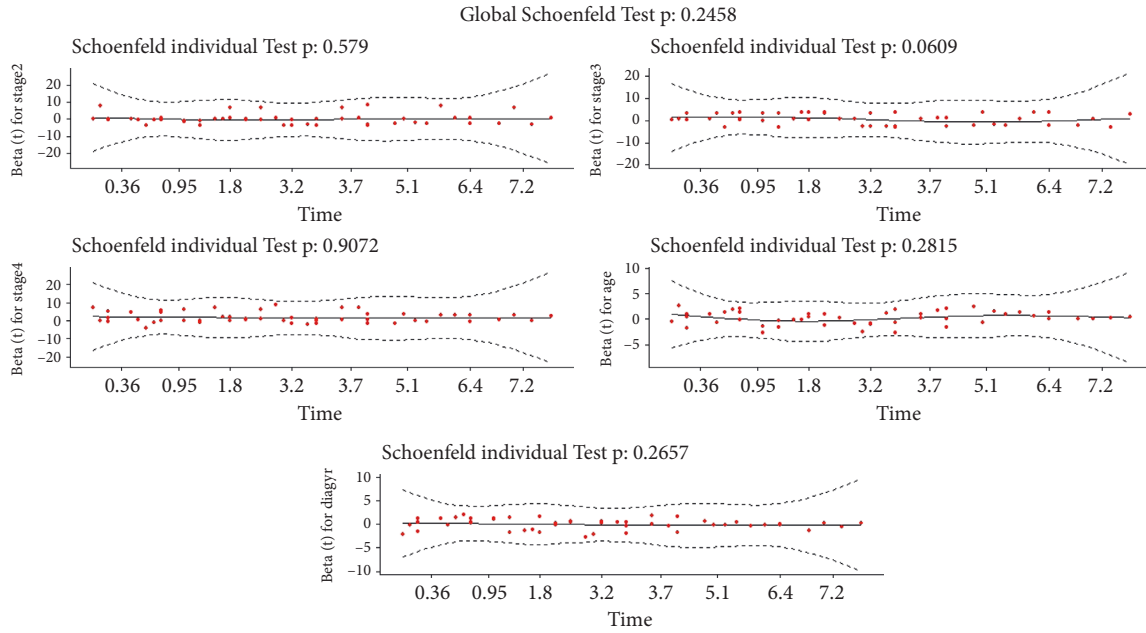
FIGURE 14: The standardized Schoenfeld residuals from the data |—larynx cancer data set, taking the test p value for each covariate into account.

TABLE 8: Numerical summaries of posterior characteristics based on McMC sample for GLL PH model for the larynx cancer data.

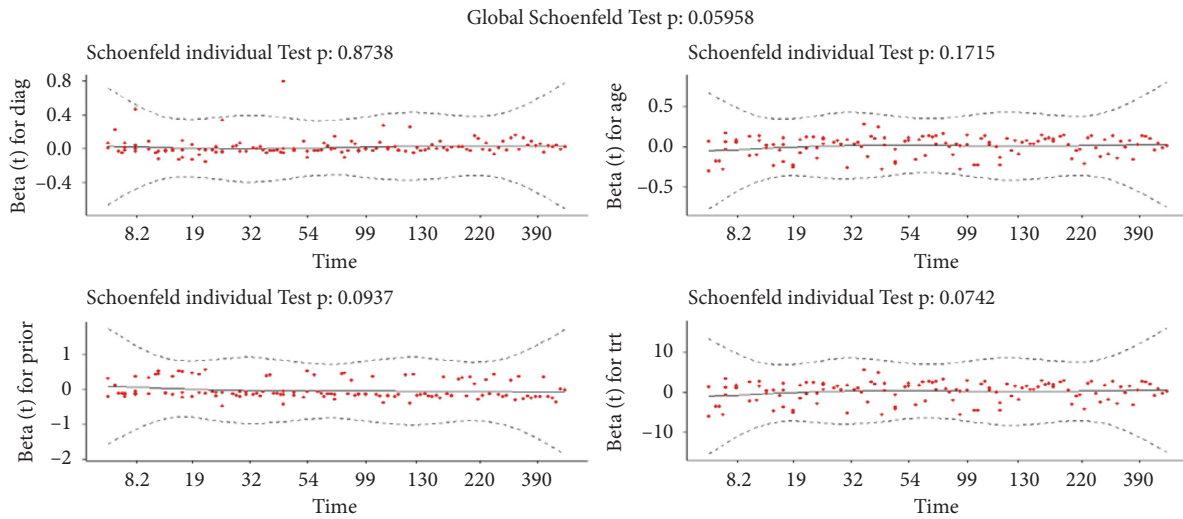| Characteristics	Pars							
	Alpha	β_1 (stage 2)	β_2 (stage 3)	β_3 (stage 4)	β_4 (age)	β_5 (diagyr)	Eta	Kappa
Mean	1.539	−0.182	0.376	1.222	0.187	−0.111	0.869	0.336
SD	0.215	0.454	0.337	0.411	0.144	0.149	0.247	0.077
Naïve SE	0.002	0.004	0.003	0.004	0.001	0.001	0.002	0.001
Time series SE	0.002	0.004	0.003	0.004	0.001	0.001	0.003	0.001
Minimum	0.847	−1.975	−0.902	−0.531	−0.373	−0.730	0.197	0.112
2.5[th] percentile	1.157	−1.108	−0.289	0.396	−0.091	−0.403	0.457	0.207
Q1	1.389	−0.480	0.152	0.952	0.089	−0.212	0.691	0.282
Medium (Q2)	1.524	−0.170	0.377	1.230	0.187	−0.112	0.846	0.328
Q3	1.668	0.128	0.605	1.498	0.284	−0.012	1.020	0.382
97.5[th] percentile	2.005	0.667	1.030	2.010	0.476	0.181	1.412	0.507
Maximum	2.701	1.648	1.770	2.848	0.817	0.509	2.131	0.763
Mode	1.550	−0.100	0.300	1.300	0.150	−0.150	0.850	0.325
Variance	0.046	0.207	0.113	0.169	0.021	0.022	0.061	0.006
Skewness	0.447	−0.173	−0.041	−0.086	0.081	0.023	0.595	0.604
Kurtosis	0.511	0.068	0.010	0.070	0.102	0.027	0.514	0.656
95% credible interval	(1.157, 2.005)	(−1.108, 0.667)	(−0.289, 1.030)	(0.396, 2.010)	(−0.091, 0.476)	(−0.730, 0.181)	(0.197, 1.412)	(0.112, 0.507)
P (>0|data)	1.000	0.352	0.870	0.998	0.906	0.227	1.000	1.000

McMC sample of posterior properties for the generalized log-logistic proportional hazard model considering the larynx data in this section.

The posterior summaries for the GLL-PH model parameters using larynx cancer data are illustrated in Table 8. The probability that the corresponding parameter is +ve is given in the last row of Table 8.

(2) Visual Summary. We looked at density strip plots (Figure 15), trace plots (Figure 16), Ergodic mean plots (Figure 17), autocorrelation plots (Figure 18), and Gelman–Rubin diagnostic plots (Figure 19), in this section, to get a visual description of the posterior properties. These plots and graphs provide a nearly comprehensive representation of the parameters' posterior uncertainty.

8.2.5. Convergence Diagnostic Tests for the Larynx Cancer Data Using GLL PH Model

(1) Statistical Tests. Table 9 indicates the Geweke, Rafter-y–Lewis, and Heidelberger–Welch diagnostics for the GLL PH model parameters.
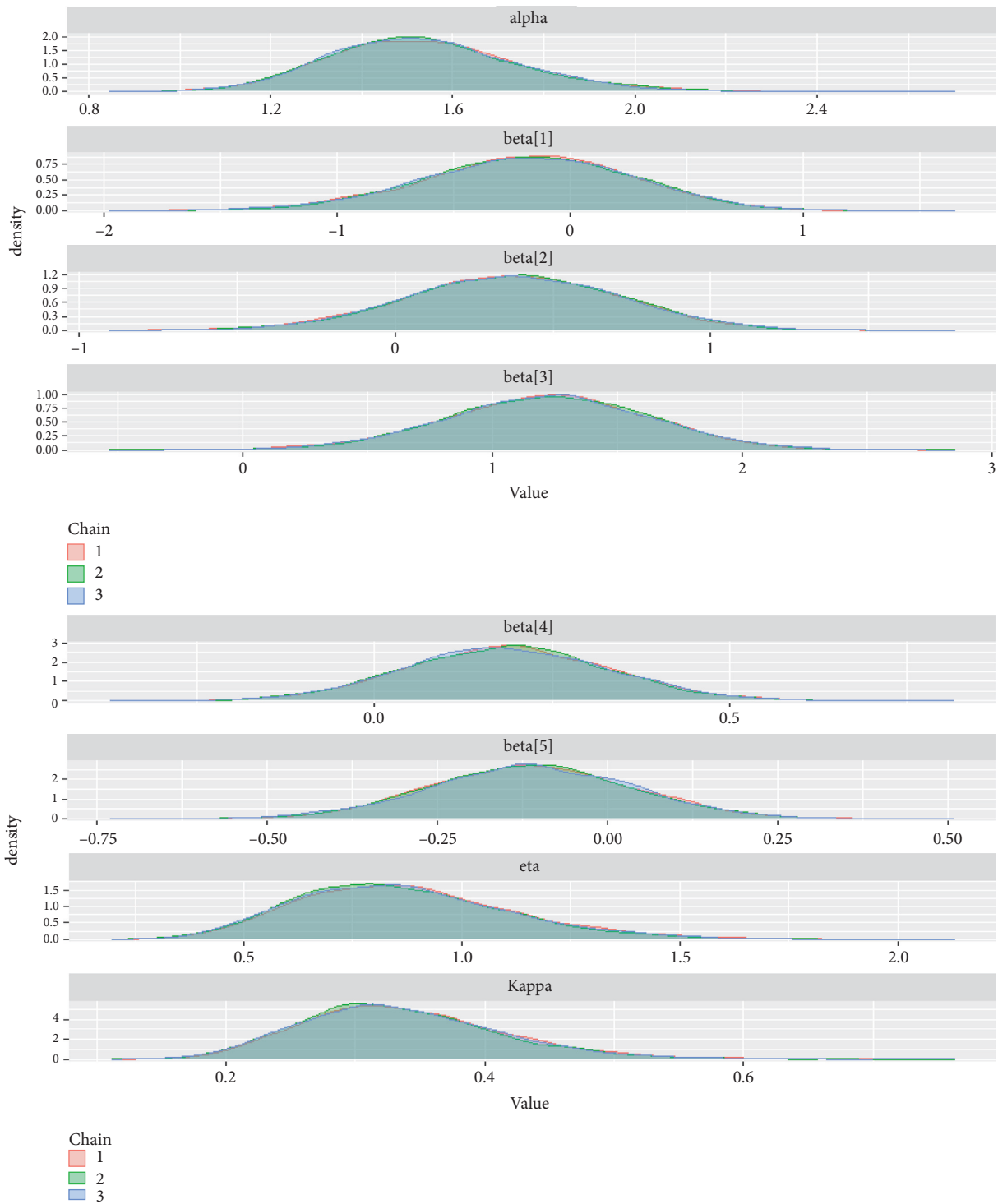
FIGURE 15: Density plots for the baseline hazard parameters and the regression coefficients for the larynx cancer data.

(2) Graphical Techniques. Convergence diagnostics of an McMC algorithm for the larynx cancer data set are presented in Figures 16–19.

8.2.6. Hazard Ratio (HR). One of the most intriguing aspects of PH models is that the regression coefficients can be interpreted using the hazard ratio, which is preferred by
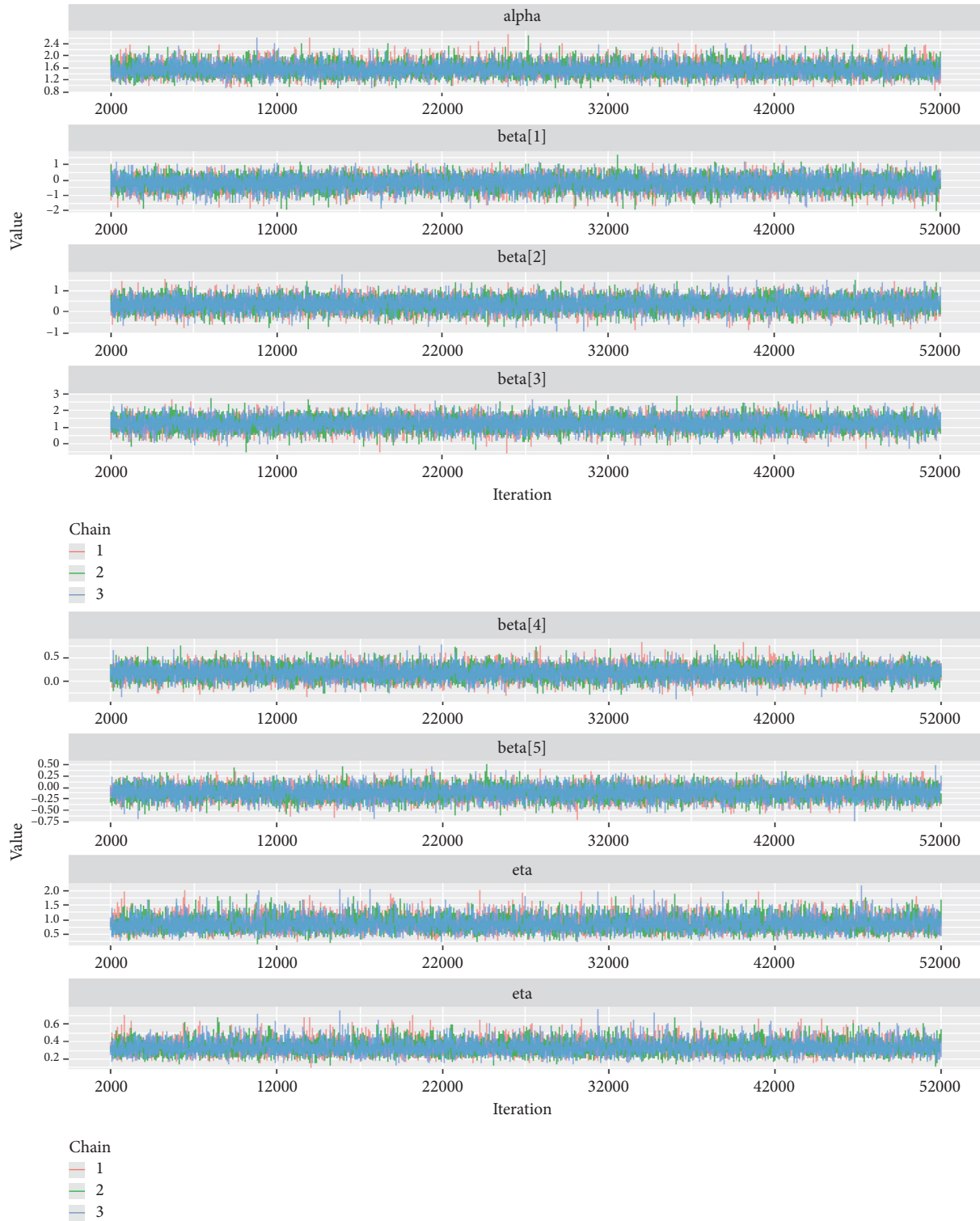
Figure 16: The time series plots for baseline hazard parameters and the regression coefficients for the larynx cancer.

many clinicians.

A key feature for PH models is the hazard ratio (HR), also known as the relative risk, between two individuals with covariate vectors $\mathbf{x}_1$ and $\mathbf{x}_2$. The HR is defined as

$$HR(\mathbf{x}_1, \mathbf{x}_2, h_0, \boldsymbol{\beta}) = \frac{h(t|\mathbf{x}_1, h_0, \boldsymbol{\beta})}{h(t|\mathbf{x}_2, h_0, \boldsymbol{\beta})} = \exp\left[(\mathbf{x}_1 - \mathbf{x}_2)^T \boldsymbol{\beta}\right], \quad (63)$$
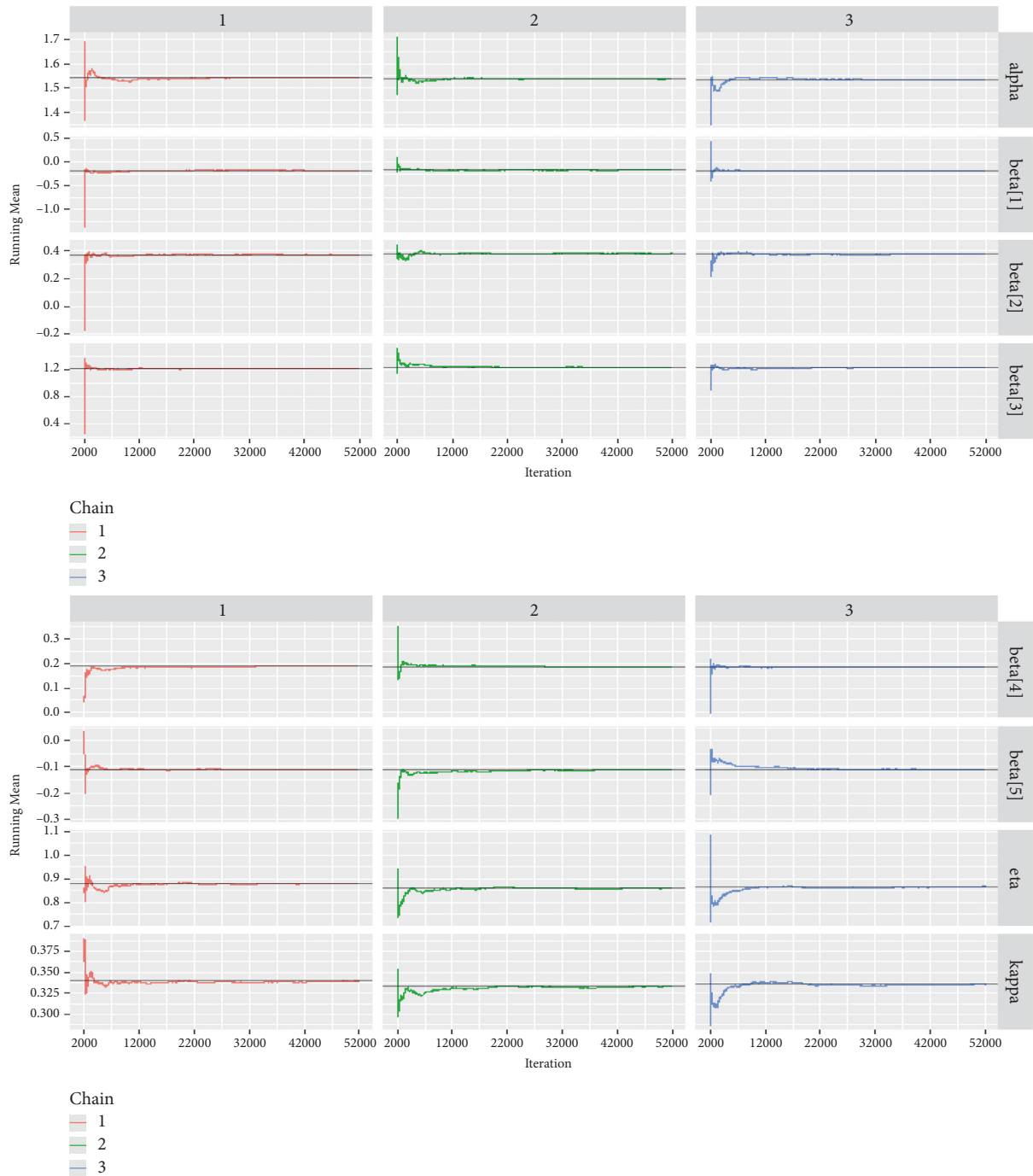
FIGURE 17: The Ergodic mean plots for the baseline hazard parameters and regression coefficients for the larynx cancer data.

which does not depend on time t. the hazard function in the numerator is equal to this constant HR times the hazard in the denominator, i.e.,

$$h\left(t|\mathbf{x}_1, h_0, \boldsymbol{\beta}\right) = HR \, x \, h\left(t|\mathbf{x}_2, h_0, \boldsymbol{\beta}\right). \tag{64}$$

Hence, the name "proportional hazards model" [19]. For example, the posterior distributions of the HR between two individuals of the same age and diagyr (year of diagnosis) but in different stages can be easily summarized.

Table 10 depicts the posterior characteristics of the hazard ratio between two men of the same age and diagnosis year (diagyr) but in different stages.

9. Bayesian Model Selection

In this study, we will use the deviance information criterion (DIC) to distinguish between the proposed models. DIC is a popular Bayesian model selection criterion. This criterion is available in most McMC packages [39]. The DIC is com-

Figure 18: Autocorrelation plots for all the baseline hazard parameters and regression coefficients for the larynx cancer data.
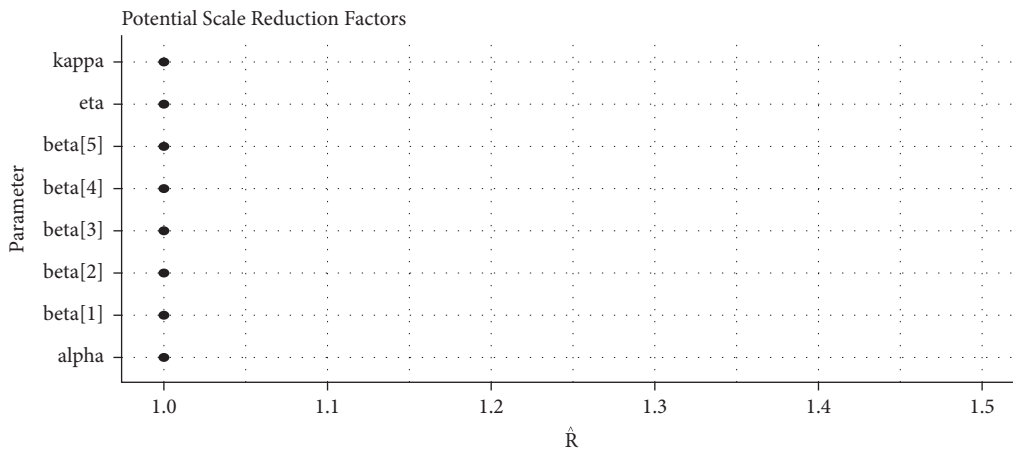
Figure 19: PSRF of the baseline hazard parameters and the regression coefficients for the larynx cancer data.

TABLE 9: Summaries for the Raftery–Lewis's, Geweke, and Heidelberger–Welch diagnostics test for the GLL PH model parameters for the right-censored larynx cancer data.

Parameter	Geweke diagnostic	Diagnostics for Raftery–Lewis	Diagnostics for the Heidelberger–Welch		
	$\Pr > \lvert z \rvert$	Dependency factor (I)	p value	Stationarity test	Half width test
Alpha	1.083	1.020	0.787	Passed	Passed
β_1 (stage 2)	−1.105	0.982	0.730	Passed	Passed
β_2 (stage 3)	−0.333	1.060	0.497	Passed	Passed
β_3 (stage 4)	0.969	1.030	0.053	Passed	Passed
β_4 (age)	−0.800	1.020	0.680	Passed	Passed
β_5 (diagyr)	−1.177	0.998	0.425	Passed	Passed
Eta	0.133	1.090	0.252	Passed	Passed
Kappa	0.317	1.150	0.189	Passed	Passed

TABLE 10: Posterior characteristics of the hazard ratio between two men of the same age and diagnosis year but in different tumour stages.

Posterior characteristics	Stages 3 and 4	Stages 2 and 4	Stages 2 and 3
Mean	0.467	0.280	0.638
Standard deviation (SD)	0.203	0.149	0.298
Naïve SE	0.001	0.001	0.002
Time series SE	0.002	0.001	0.002
2.5%	0.197	0.088	0.218
Lower quartile (Q1)	0.326	0.175	0.423
Medium (Q2)	0.425	0.250	0.585
Upper quartile (Q3)	0.564	0.349	0.791
97.5%	0.967	0.648	1.366

TABLE 11: Posterior properties summaries and the information criterion values for the considered GLL PH model and its competing models for the lung cancer data.

Summaries	Posterior characteristics				
Parametric competitive models	Parameter(s)	Posterior mean	Posterior SD	$\Pr (> \lvert 0 \rvert \text{data})$	HPD interval (95%)
GLL-PH model (DIC = 1505.165)					
	Alpha	1.317	0.173	1.000	(1.001, 1.661)
	β_1 (diagt)	0.002	0.010	0.598	(−0.019, 0.021)
	β_2 (age)	−0.024	0.008	0.003	(−0.039, −0.007)
	β_3 (prior)	−0.015	0.021	0.244	(−0.057, 0.026)
	β_4 (trt)	−0.151	0.178	0.199	(−0.505, 0.186)
	Eta	0.042	0.015	1.000	(0.016, 0.073)
	Kappa	0.103	0.049	1.000	(0.029, 0.200)
Weibull-PH model (DIC = 1521.310)					
	Alpha	0.744	0.048	1.000	(0.654, 0.842)
	β_1 (diagt)	0.005	0.010	0.648	(−0.018, 0.024)
	β_2 (age)	−0.025	0.007	0.001	(−0.039, −0.010)
	β_3 (prior)	−1.027	0.021	0.102	(−0.068, 0.015)
	β_4 (trt)	−0.252	0.180	0.080	(−0.593, 0.108)
	Kappa	0.206	0.090	1.000	(0.060, 0.388)
Gompertz-PH model (DIC = 1556.407)					
	Alpha	1.134	0.311	1.000	(0.567, 1.746)
	β_1 (diagt)	0.021	0.009	0.984	(0.003, 0.039)
	β_2 (age)	0.027	0.006	1.000	(0.014, 0.039)
	β_3 (prior)	−0.056	0.023	0.006	(−0.099, −0.012)
	β_4 (trt)	−0.136	0.182	0.228	(−0.494, −0.211)
	Kappa	0.001	0.0002	1.000	(0.001, 0.002)

TABLE 12: Posterior properties summaries and the information criterion values for the considered GLL PH model and its competing models for the larynx cancer data.

Summaries		Posterior characteristics			
Parametric competitive models	Parameter(s)	Posterior mean	Posterior SD	Pr (>|0|data)	HPD interval (95%)
GLL-PH model (DIC = 294.412)					
	Alpha	1.539	0.215	1.000	(1.157, 2.005)
	β_1 (stage 2)	−0.182	0.454	0.352	(−1.108, 0.667)
	β_2 (stage 3)	0.376	0.337	0.870	(−0.289, 1.030)
	β_3 (stage 4)	1.222	0.411	0.998	(0.396, 2.010)
	β_4 (age)	0.187	0.144	0.906	(−0.091, 0.476)
	β_5 (diagyr)	−0.111	0.149	0.227	−0.730, 0.181)
	Eta	0.869	0.247	1.000	(0.197, 1.412)
	Kappa	0.336	0.077	1.000	(0.112, 0.507)
Weibull-PH model (DIC = 296.776)					
	Alpha	0.908	0.105	1.000	(0.713, 1.118)
	β_1 (stage 2)	−0.380	0.446	0.198	(−1.275, 0.468)
	β_2 (stage 3)	0.174	0.318	0.711	(−0.483, 0.781)
	β_3 (stage 4)	1.095	0.393	0.997	(0.329, 1.857)
	β_4 (age)	0.176	0.141	0.899	(−0.092, 0.461)
	β_5 (diagyr)	−0.012	0.146	0.468	(−0.294, 0.274)
	Kappa	0.154	0.041	1.000	(0.081, 0.236)
Gompertz-PH model (WAIC = 297.560)					
	Alpha	0.134	0.031	1.000	(0.076, 0.196)
	β_1 (stage 2)	−0.138	0.455	0.392	(−1.040, 0.737)
	β_2 (stage 3)	0.393	0.328	0.886	(−0.252, 1.041)
	β_3 (stage 4)	1.544	0.397	1.000	(0.776, 2.308)
	β_4 (age)	0.206	0.149	0.919	(−0.084, −0.501)
	β_5 (diagyr)	0.075	0.155	0.685	(−0.230, 0.374)
	Kappa	0.552	0.186	1.000	(0.227, 0.919)

puted as follows:

$$DIC = \overline{D} + pD = \widehat{D} + 2pD, \qquad (65)$$

where $\overline{D}$ denotes the deviance's posterior mean and is a goodness of fit test for parametric survival models and pD calculates as the difference between $pD = \overline{D} - \widehat{D}$, and it is denoted the effective number of proposed model parameters.

9.1. Data Set I. Table 11 displays some posterior characteristics for the three PH models (generalized log-logistic, Gompertz, and Weibull). Even though the estimates of the regression coefficient are significant compared, the flexibility provided by the GLL distribution's additional shape parameter contributes to its ultimate superiority over the Gompertz and Weibull models and the DIC shows us its goodness-of-fit and versatility comparing to the competing parametric PH models.

9.2. Data Set II. Table 12 displays some posterior characteristics for the three PH models (generalized log-logistic, Gompertz, and Weibull). Even though the estimates of the regression coefficient are significant compared, the flexibility provided by the GLL distribution's additional shape parameter contributes to its ultimate superiority over the Gompertz and Weibull models and the DIC demonstrates us

its goodness-of-fit and versatility comparing to the competing parametric PH models.

10. Conclusion and Future Work

In this paper, we explored how to derive Bayesian estimates of the baseline hazard parameters and the regression coefficients of the parametric proportional hazard model with generalized log-logistic baseline hazard using right-censored survival data utilizing McMC approaches. The McMC techniques offer an alternative technique for estimating the parameters of the proposed model that is more flexible than frequentist techniques such as maximum likelihood estimation. Bayesian inference was performed with a variety of priors, and the convergence pattern was investigated using various diagnostic procedures.

To test the performance of the proposed parametric PH model, a comprehensive McMC simulation study was conducted. According to the simulation results, the PH model produces better results, with fewer absolute biases and MSEs for most regression coefficients and baseline distributional parameters. The behavior of the PH model in a generic PH regression situation comprising numerous covariates was also examined using synthetic right-censored data sets. Our findings indicate that the PH model performs well when handling with multiple factors. The paper's final analysis focused on a real-world application involving two well-known right-censored survival data sets for lung cancer

and laryngeal cancer patients. In conclusion, the findings of the proposed parametric PH model show that it performs better and is superior to the other competing PH model, as well as indicating significant distributional parameters and regression coefficients.

Furthermore, for both simulation and real-data analysis, the regression coefficients were assumed to have a normal prior, and the baseline distribution parameters were assumed to have an independent gamma prior to compute the quantities of importance derived from the proposed model's posterior distribution. It has been attempted to create a visual summary and other essential graphs to aid in the interpretation of results and decision making. Finally, we hope that this paper will be an extension of the work of Khan and Khosa [11] and will encourage researchers who employ parametric hazard-based regression models to conduct their analyses using the Bayesian approach from the BUGs codes with the help of the R software's RJAGS package.

In terms of future work, we intend to produce an R package to fit the most prevalent parametric hazard-based regression models, including the PH model. The method given in this study can also be applied to multiple event scenarios, such as the competing risk model, and to survival data with a cure fraction rate. It can also be applied to joint model frameworks. Other types of censored and truncated observations, such as left censoring, interval censoring, and double censoring, could be used in future research. This is outside the scope of this study and will be addressed in future ones.

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.

Authors' Contributions

All authors contribute significantly to this manuscript.

Acknowledgments

References

[1] D. Alvares, E. Lázaro, V. Gómez-Rubio, and C. Armero, "Bayesian survival analysis with BUGS," *Statistics in Medicine*, vol. 40, no. 12, pp. 2975–3020, 2021.

[2] R. Alanazi, "Identification and prediction of chronic diseases using machine learning approach," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–9, Article ID 2826127, 2022.

[3] R. M. K. Mohamed, O. R. Shahin, N. O. Hamed, H. Y. Zahran, and M. H. Abdellattif, "Analyzing the patient behavior for improving the medical treatment using smart healthcare and IoT-based deep belief network," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–8, Article ID 6389069, 2022.

[4] C. Lee, W. Zame, J. Yoon, and M. Van Der Schaar, "Deephit: a deep learning approach to survival analysis with competing risks," in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, vol. 32, New Orleans, LA, USA, February 2018.

[5] M. Elhoseny, Z. Tarek, and I. M. El-Hasnony, "Advanced cognitive algorithm for biomedical data processing: COVID-19 pattern recognition as a case study," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–11, Article ID 1773259, 2022.

[6] D. Collett, *Modelling Survival Data in Medical Research*, CRC Press, Florida, USA, 2015.

[7] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New Jersey, USA, 2011.

[8] X. Wang, Y. Yue, and J. J. Faraway, *Bayesian Regression Modeling with INLA*, Chapman and Hall/CRC, London, UK, 2018.

[9] F. J. Rubio, L. Remontet, N. P. Jewell, and A. Belot, "On a general structure for hazard-based regression models: an application to population-based cancer research," *Statistical Methods in Medical Research*, vol. 28, no. 8, pp. 2404–2417, 2019.

[10] C. Legrand, *Advanced Survival Models*, Chapman and Hall/CRC, London, UK, 2021.

[11] S. A. Khan and S. K. Khosa, "Generalized log-logistic proportional hazard model with applications in survival analysis," *Journal of Statistical Distributions and Applications*, vol. 3, no. 1, p. 16, 2016.

[12] A. H. Muse, S. M. Mwalili, and O. Ngesa, "On the log-logistic distribution and its generalizations: a survey," *International Journal of Statistics and Probability*, vol. 10, no. 3, p. 93, 2021.

[13] A. H. Muse, A. H. Tolba, E. Fayad, O. A. Abu Ali, M. Nagy, and M. Yusuf, "Modelling the COVID-19 mortality rate with a new versatile modification of the log-logistic distribution," *Computational Intelligence and Neuroscience*, vol. 2021, no. November, pp. 1–14, 2021, Nov. 2021.

[14] G. M. Cordeiro, R. B. Silva, and A. D. C. Nascimento, *Recent Advances in Lifetime and Reliability Models*, Bentham, London, UK, 2020.

[15] M. H. Tahir and G. M. Cordeiro, "Compounding of distributions: a survey and new generalized classes," *Journal of Statistical Distributions and Applications*, vol. 3, no. 1, 2016.

[16] M. H. Tahir and S. Nadarajah, "Parameter induction in continuous univariate distributions: well-established G families," *Anais da Academia Brasileira de Ciências*, vol. 87, no. 2, pp. 539–568, 2015.

[17] A. H. Muse, S. Mwalili, O. Ngesa, S. J. Almalki, and G. A. Abd-Elmougod, "Bayesian and classical inference for the generalized log-logistic distribution with applications to survival data," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–24, Article ID 5820435, 2021.

[18] S. A. Khan and N. Basharat, "Accelerated failure time models for recurrent event data analysis and joint modeling," *Computational Statistics*, pp. 1–29, 2021.

[19] R. Christensen, W. Johnson, A. Branscum, and T. E. Hanson, *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*, CRC Press, Florida, USA, 2010.

[20] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B*, vol. 34, no. 2, pp. 187–202, 1972.

[21] I. Alkhairy, M. Nagy, A. H. Muse, and E. Hussam, "The arctan-X family of distributions: properties, simulation, and applications to actuarial sciences," *Complexity*, vol. 2021, pp. 1–14, Article ID 4689010, 2021.

[22] A. H. Muse, S. Mwalili, O. Ngesa, H. M. Alshanbari, S. K. Khosa, and E. Hussam, "Bayesian and frequentist approach for the generalized log-logistic accelerated failure time model with applications to larynx-cancer patients," *Alexandria Engineering Journal*, vol. 61, no. 10, pp. 7953–7978, 2022.

[23] M. Y. Danish and M. Aslam, "Bayesian estimation in the proportional hazards model of random censorship under asymmetric loss functions," *Data Science Journal*, vol. 11, no. November, pp. 72–88, 2012.

[24] M. Y. Danish and M. Aslam, "Bayesian estimation in random censorship model for Weibull distribution under different loss functions," *Advances in Adaptive Data Analysis*, vol. 04, no. 03, Article ID 1250021, 2012.

[25] F. J. Rubio, B. Rachet, R. Giorgi, C. Maringe, and A. Belot, "On models for the estimation of the excess mortality hazard in case of insufficiently stratified life tables," *Biostatistics*, vol. 22, no. 1, pp. 51–67, 2021.

[26] R. C. TeamR. A. Lang Environ. Stat. Comput. R Found. Stat. Comput. Vienna, 2020.

[27] M. Plummer, *JAGS: A Program for Analysis of Bayesian Graphical Modeling Using Gibbs Sampling [computer Program]*, 2017.

[28] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and B. D. Rubin, *Bayesian Data Analysis*, Chapman and Hall/ CRC, London, UK, 3rd editio edition, 2013.

[29] J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, Vol. 360, John Wiley & Sons, , New Jersey, USA, 2011.

[30] C. G. B. Balieiro, *KGsurv: an R package to fit the Kumaraswamy-G family of distributions for survival data*, 2021.

[31] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, no. 4, pp. 457–472, 1992.

[32] B. J. Smith, *Bayesian Output Analysis Program (BOA) Version 1.1 User's Manual*, Dept. Biostat. Univ. Iowa, Coll. Public Heal, Lowa, 2005.

[33] S. Sinharay, "Assessing convergence of the Markov chain Monte Carlo algorithms: a review," *ETS Research Report Series*, vol. 2003, no. 1, 52 pages, 2003.

[34] V. Roy, "Convergence diagnostics for Markov chain Monte Carlo," *Annual Review of Statistics and Its Application*, vol. 7, no. 1, pp. 387–412, 2020.

[35] A. E. Raftery and S. M. Lewis, *Implementing Mcmc,*" Markov Chain Monte Carlo Pract, pp. 115–130, 1996.

[36] P. Heidelberger and P. D. Welch, "Simulation run length control in the presence of an initial transient," *Operations Research*, vol. 31, no. 6, pp. 1109–1144, 1983.

[37] J. Geweke, "Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments," *Bayesian Stat*, vol. 4, pp. 641–649, 1992.

[38] O. Kardaun, "Statistical survival analysis of male larynx-cancer patients - a case study," *Statistica Neerlandica*, vol. 37, no. 3, pp. 103–125, 1983.

[39] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde, "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B*, vol. 64, no. 4, pp. 583–639, 2002.

[40] P. C. Austin, "Generating survival times to simulate Cox proportional hazards models with time-varying covariates," *Statistics in Medicine*, vol. 31, no. 29, pp. 3946–3958, 2012.

[41] R. Bender, T. Augustin, and M. Blettner, "Generating survival times to simulate Cox proportional hazards models," *Statistics in Medicine*, vol. 24, no. 11, pp. 1713–1723, 2005.

[42] A. E. Raftery and S. M. Lewis, "The number of iterations, convergence diagnostics and generic Metropolis algorithms," *Pract. Markov Chain Monte Carlo*, vol. 7, no. 98, pp. 763–773, 1995.

[43] A. E. Raftery and S. M. Lewis, "[Practical Markov chain Monte Carlo]: comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo," *Statistical Science*, vol. 7, no. 4, pp. 493–497, 1992.

Research Article

Predicting the Kidney Graft Survival Using Optimized African Buffalo-Based Artificial Neural Network

Riddhi Chawla [iD],[1] **S. Balaji,**[2] **Raed N. Alabdali,**[3] **Ibrahim A. Naguib** [iD],[4] **Nadir O. Hamed** [iD],[5] **and Heba Y. Zahran**[6,7,8]

[1]*Medical School, Akfa University, Tashkent, Uzbekistan*
[2]*Department of Computer Science Engineering, Panimalar Engineering College, Chennai, Tamil Nadu, India*
[3]*Department of Computer Science, College of Science and Arts in Qurayyat, Jouf University, Sakakah, Saudi Arabia*
[4]*Department of Pharmaceutical Chemistry, College of Pharmacy, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia*
[5]*Computer Studies Department, Elgraif Sharg Technological College, Sudan Technological University, Khartoum, Sudan*
[6]*Laboratory of Nano-Smart Materials for Science and Technology (LNSMST), Department of Physics, Faculty of Science, King Khalid University, Abha 61413, Saudi Arabia*
[7]*Research Center for Advanced Materials Science (RCAMS), King Khalid University, Abha 61413, Saudi Arabia*
[8]*Nanoscience Laboratory for Environmental and Biomedical Applications (NLEBA), Metallurgical Laboratory 1, Department of Physics, Faculty of Education, Ain Shams University, Roxy, Cairo 11757, Egypt*

Correspondence should be addressed to Riddhi Chawla; r.chawla@akfauniversity.org and
Nadir O. Hamed; nohamedit@gmail.com

A variety of receptor and donor characteristics influence long-and short-term kidney graft survival. It is critical to predict the effectiveness of kidney transplantation to optimise organ allocation. This would allow patients to choose the best accessible kidney donor and the optimal immunosuppressive medication. Several studies have attempted to identify factors that predispose to graft rejection, but the results have been contradictory. As a result, the goal of this paper is to use the African buffalo-based artificial neural network (AB-ANN) approach to uncover predictive risk variables related to kidney graft. These two feature selection approaches combine to provide a novel hybrid feature selection technique that could select the most important elements to improve prediction accuracy. The feature analysis revealed that clinical features have varied effects on transplant survival. The collected data is processed in both training and testing methods. The prediction model's performance, in terms of accuracy, precision, recall, and F-measure, was examined, and the results were compared with those of other existing systems, including naive Bayesian, random forest, and J48 classifier. The results suggest that the proposed approach can forecast graft survival in kidney recipients' next visits in a creative manner and with more accuracy compared with other classifiers. This proposed method is more efficient for predicting kidney graft survival. Incorporating those clinical tools into outpatient clinics' everyday workflows could help physicians make better and more personalised decisions.

1. Introduction

The importance of predicting the outcome of kidney transplantation cannot be overstated [1]. Research scholars and decision makers are progressively being urged to promote patient-centred care that respects the preferences, requirements, and values of patients. Patients with end-stage organ dysfunction require organ transplant, which improves their quality of life [2]. The capacity to forecast survival rate

after transplant is vital and plays a key role in comprehending the donor-recipient matching procedure. This matching is essential for renal replacement success because it allows patients to choose the fine accessible kidney donor and the finest immunosuppressive medication. Prognosis of organ transplantation outcome is a clinically important and difficult subject. Predicting survival before treatment simplifies the patient's decisions and improves survival by influencing clinical practise decisions [3]. Many variables

that influence the prediction problems have been extensively investigated, but the complicated relationship among these variables make prediction process a difficult task. Kidney transplantation is regarded as the potential alternate medication for individuals having end-stage renal illness since it has several benefits over dialysis, including a higher quality of life and a longer survival rate [4].

Graft functioning and survivals have improved significantly over the last two decades, yet several transplanted kidneys are discarded due to chronic allograft nephropathy and acute rejection [5]. Compared with the individuals with functioning grafts, this results in a three-fold increased risk of death. In terms of results, it has long been suspected that in the case of kidney transplantation, patient preferences prefer graft survival over the danger of illness or malignancy. Prediction of individual graft survival [6] could thus be an initial step in enhancing patient's health status information and promote patient-centred care. Because of the scarcity of organs, long waiting lists, the higher retransplantation costs, the risk of graft failure, and kidney graft performance must be closely monitored. A variety of receptor-donor related parameters that affect graft survival affect the kidney transplant distribution. As the demand for kidney transplantation grows around the world, it is essential to recognize the possible issues for graft failure so as to enhance the survivability of patients and the quality of their life [7]. Investigating, identifying, and adjusting for risk variables are critical because transplantation failure is connected with negative outcomes for patients. Nevertheless, due to the obvious wide range of risk variables for graft failure, this evidence is much harder to quantify at an individual scale [8].

Various prognostic and predictive factors impacting the effectiveness of renal grafts were explored in different researches, including age of donor and receptor, sex, type of donor (alive or deceased), body mass index, anaemia, kind of immunosuppressive regimen, and so on. However, the outcomes were contradictory. Several clinical investigations on the impact of these parameters on graft survival have been undertaken [9], but considering the complicated interplay among those factors, still there is more to explore in this domain. With receiver operating characteristic scores, current risk forecasting models could only predict the outcomes of kidney transplantation recipients to a smaller extent. On the basis of covariates and predictors, numerous classification techniques are employed for predicting a categorical response variable. Although neural networks may predict the whole clinical results, they cannot discern particular risk variables for a specific clinical event [10]. The existence of unrelated factors may increase the approach's difficulty, making it hard to build a predictive model utilizing clinical data.

Machine learning approaches presented in this field have shown a reliable and robust performance in categorizing dualistic responses. To develop nonlinear models, the artificial neural network strategy is introduced, and it is capable of automatically detecting complicated nonlinear correlations among dependent and independent variables, as well as all conceivable relations among predictor

variables [11]. Kidney transplant is the most effective therapy for end-stage kidney problems. It enhances survival of patients and provides a greater quality of life than haemodialysis. Moreover, it decreases the long-term healthcare costs for such individuals significantly. Extended immunosuppression, on the other hand, is connected to a number of adverse effects that could change both patient and graft survivability. Graft survival is the period of time that a kidney transplant (graft) works well enough for the patient not to require dialysis or any other transplant method. The goal of the research is to establish a novel forecasting approach which combines feature engineering with the deep learning techniques via an optimization mechanism in order to increase prediction performance. To accomplish this goal, a unique prediction approach based on kidney graft survival data has been developed for forecasting the survival of graft after transplantation, which could be used in real-time and is suitable for forecasting kidney transplantation outcomes via data analysis. Any transplantation dataset can be used with the proposed prediction algorithm.

The remainder of the article is laid out as follows. The present researches on the prediction of kidney transplantation graft survival are examined in Section 2. The novel proposed AB-ANN prediction approach is presented in detailed manner in Section 3. Section 4 discusses the included dataset, as well as the test results, and Section 5 describes the discussions, and Section 6 concludes this study.

2. Related Works

2.1. Predictive Modelling Technique. Data-driven strategies were used in a number of studies to predict graft survival following transplantation. The authors in [12] investigated the factors impacting graft survival before and after kidney transplantation by employing Kaplan–Meier methods. To improve organ retrieval allocation, a multivariate analysis [13] was utilized for predicting the kidney transplantation outcomes using a deceased donor. However, by relying solely on statistical methodologies, these researches were limited. As a result, better methodologies are needed to uncover potentially hidden information among the various characteristics that could influence the graft survival state forecasting of a kidney transplant.

To determine graft survival out of a deceased individual, a tree regression model [14] has been introduced. After transplanting kidney, a neural network strategy was developed for estimating the delayed graft functioning [3]. However, those researches were limited to the deceased donors. Other researchers sought to develop transplantation outcome prediction models. To identify essential variables and subsequently design a Bayesian belief network, researchers utilized statistical mechanisms like elastic nets with machine learning techniques like ANN, bootstrap, random forest, and support vector machines. This model looked into the variables' hidden dependencies. This model had a precision of 68.4%.

Cox-based models [15] have been used extensively in the survival assessment of complex organ transplants; but when

the feature space grows larger, such techniques lose the prediction accuracy. A feature selection scheme based on a hybrid genetic algorithm determines the key traits for lung transplantation. They employed three different classification prediction processes for forecasting the lung transplantation and quality of life of patients [16]. The findings excelled the previous research. Some other studies deployed artificial neural networks and a statistically determined nomogram for forecasting the five-year graft survival after transplanting the kidney, using clinical and demographic data [17]. Using an external validation dataset, they discovered that the artificial neural networks outperformed the nomograms. The authors in [18] created a Bayesian belief network for predicting the graft survival. With good accuracy, the model can determine the graft failure. A Bayesian belief network architecture is employed in some other studies for forecasting the heart transplantation outcome. Compared with other approaches in the literature, the results showed identical predictive effectiveness.

In kidney transplantation, machine learning-based predictive algorithms identify the main correlations among receptor and donor characteristics for predicting transplant outcomes based upon acceptor-donor data. ML approaches were used in a number of researches for predicting the outcome of kidney graft [19], but almost in all examined studies, the conventional mechanism has been to choose one or more arbitrary time periods commencing from the transplant date and use categorization techniques for predictive purpose. In terms of prediction modelling and feature engineering, there is a definite requirement for more research into data stratification methodologies and other machine learning methods [20].

2.2. Explanatory Modelling Technique. The alternative models for kidney graft and receiver survival prediction include artificial neural networks as well as linear regression mechanisms [21]. Other approaches like landmark modelling and joint modelling utilize time-dependent factors to increase predictive performance in addition to such approaches that have used static covariates. The feature selection is a key issue in a variety of fields, including document classification, prediction object identification, and bioinformatics, as well as the representation of complicated production technologies. In such applications, datasets with hundreds of features are frequent. For some situations, all the features could be significant, but for certain target concepts, just a small subset of features are highly essential. Some classification techniques have learned to focus on the most critical features while ignoring the less important feature points. Decision trees are one type of such methods; however, multilayer perceptron neural networks with significant normalization of the input layer also can automatically eliminate unnecessary features [22].

The kidney graft survival is derived by Bayesian belief network modelling. In this research, the 5155 patients were randomly selected from the database of renal data system in US.

The key contributions of this research are as follows:

(i) Introducing a newly proposed African buffalo optimization for feature selection, which could effectively choose the most relevant feature set for prediction

(ii) Designing a newly combined predictive model, which could correctly assess the status of kidney graft transplantation and improve the limitations in the prior studies

(iii) Combining information gain function and the ABO mechanism with the ANN model to attain good predictive abilities

3. Proposed AB-ANN Methodology

The research extends to the prediction of graft survival approach by proposing a new three-phase approach, that is, (i) data processing phase, (ii) feature selection phase, and (iii) prediction phase [23].

Prior to data processing, donor and recipient characteristics such as age, gender, blood type, and health are analysed. The cross match test is used to find out how the donor's blood reacts with the recipient's blood, and the HLA test analyses the immune system to determine the outcome of the operation.

The input data is first gathered and preprocessed to be used for training and testing purpose. Data cleaning and data censoring are the two phases of data processing phase. Followed by this, feature selection is accomplished to recognize the most essential features which would be used in the prediction phase, reducing both the complexity of the technique and the features dimensionality. Information gain along with the ABO mechanism is utilized to choose the most relevant features. These two feature selection approaches combine to provide a novel hybrid feature selection technique, which could select the most important elements to improve prediction accuracy. Finally, the status of the graft is forecasted as survive or not survive in the prediction phase. The workflow of the proposed AB-ANN model is represented in Figure 1.

3.1. Training and Testing Data. The kidney transplantation dataset given by Mansoura University's Urology and Nephrology Center [24] was utilized to validate the suggested prediction approach for predicting graft survival. This database includes medical history, demographic data, some preoperative considerations for either recipient and donors, physical situations both during and after transplantation, and extra features like transplant date and dialysis information of kidney transplant patients. The data is divided into two categories: training (70%) and testing (30%). Initially, a portion of the dataset is utilized to train the proposed predictive model (training set). The system is then utilized to forecast survival class by testing a new subset of the dataset (test set).

3.2. Data Processing Phase. The input data is preprocessed during the data processing step, so that it may be used during
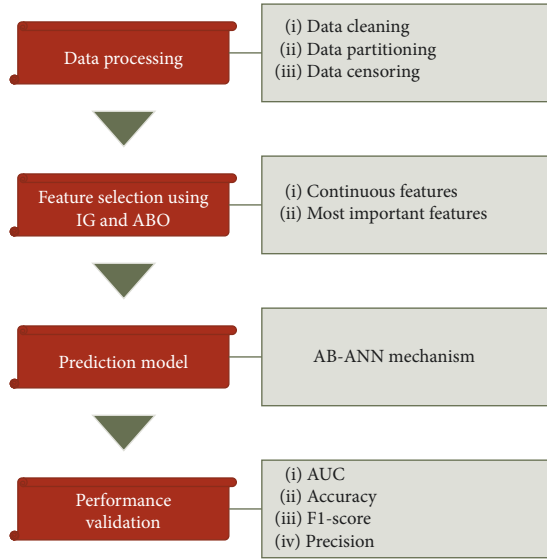
Figure 1: A novel AB-ANN prediction model.

training and testing. The dataset was preprocessed during this stage. Data cleaning and data censoring are the two phases that make up this process.

Because the prediction approach is meant to forecast the outcome of kidney transplantation before transplanting, all operational and postoperational features are deleted during the data cleaning process. The traits that will have no predictive value are removed in the second level (e.g., patient's name, the hospital ID, and date of examination). In the last stage, certain occurrences are deleted since the dataset contains missing data. Missing value imputation can be done in a variety of ways. For 1 percent missing values, the custom mean imputation approach was utilized, in which each covariate's missing values were replaced with the mean of its preceding and next values in the temporal order.

Graft survival condition was censored in the data censoring process when the graft time was less than the number of days in the five-year period and the graft was still alive, or the study finishing date. If the patient is on dialysis or died with a failing graft, the graft time is calculated by deducting the transplantation date from the dialysis initiation date. If the individual's state is surviving with functional graft or died with functioning graft, deduct the transplantation date from the last follow-up date.

3.3. Functioning of AB-ANN Mechanism

3.3.1. Feature Selection Phase. Selection of features is a significant part in any data mining procedures. Choosing the most important features would increase the prediction accuracy of the model thereby reducing the computation time and processing costs. An optimised feature selection approach called African buffalo optimization mechanism with information gain function is used in this study to successfully determine the most significant features that could improve the prediction process. The novel combined feature selection mechanism combines the advantages of either

method, resulting in significantly improved system performance.

To get started, IGBFS is structured to identify key attributes based on choosing which features to use. To use UNC databases, identify 55 features of 67 attributes as important. These important properties vary with IG rather than zero. Second, in addition to selecting the most important features, NBBFS was used to specify the most important features from the basic features developed by the IGBFS system.

The goal of employing information gain (IG) is to identify characteristics that provide the most significant knowledge about the classes [25]. Such characteristics are primarily discriminatory and occur within a single class. IG is a feature ranking methodology that utilizes entropy to calculate the degree to which the entropy is reduced when observing the value of a particular feature. As a result, the value of information gain indicates how much information this feature contributes to the database. Each feature has an information gain rating that indicates whether it is necessary or not. As a result, the feature with IG = 0 is rejected. With a higher IG, the chances of attaining clear classes in the target class increase.

The critical characteristics are determined after calculating the information gain values for all features. The qualities with an information gain value higher than zero are considered as essential. The features are examined using an edge value; if a feature's information gain value is more than the edge value, it is chosen; otherwise, it is not. In this study, a threshold of zero is employed, and features having an information gain more than zero are regarded the most essential features for prediction.

The African buffalo optimization mechanism [26] selects the essential features for prediction using its fitness function. The African buffalo optimization mechanism is also employed as an optimizer in the last layer of the AB-ANN model for enhancing the prediction accuracy. Furthermore, the learning factors help to process the trace of essential feature points. The cooperative behaviour of buffalo is reorganised by $le_1(\beta_p^{\text{targ}} - w_f)$, and the intelligence of the buffalo is denoted by $le_2(b_{p\max.f} - w_f)$. Also, the fitness value is computed by

$$m_f + 1 = m_f + le_1\left(\beta_p^{\text{targ}} - w_f\right) + le_2\left(b_{p\max.f} - w_f\right). \quad (1)$$

Here, $m_f + 1$ denotes the next feature, and also, m_f represents the current feature value. In addition, new feature update is deliberated using

$$w_f + 1 = \frac{w_f + m_f}{\lambda^*}, \quad (2)$$

where w_f and m_f indicate the respective exploration and exploitation fitness of f.

3.3.2. Prediction Phase. For predictive purpose, an artificial neural network was deployed. It assigns the tested instances to the class with the greatest likelihood. It is assumed that impact of features on a class is unaffected by other variables. The enhanced ANN model speeds up and improves the

computation accuracy. A multilayer feed forward perceptron was employed as the neural network [3]. The following (3)–(5) are the mathematical depiction of a neural network.

Let

$$t_1 = f'\left(m_1 + \mu_{11}l_1 + \mu_{12}l_2 + \cdots + \mu_{1p}l_p\right). \qquad (3)$$

$$t_n = f'\left(m_n + \mu_{n1}l_1 + \mu_{n2}l_2 + \cdots + \mu_{np}l_p\right), \qquad (4)$$

where t_n is the output from mth hidden node, m is the total number of nodes in the hidden layer, n is the number of covariate, m is the intercept parameter, l_p is the pth covariate, μ_{np} is the pth covariate and nth hidden node parameter, and $f'(\cdot)$ is considered to be the activation function.

Now,

$$s = g'\left(a_1 + b_1t_1 + \cdots + b_nt_n\right), \qquad (5)$$

where s indicates the neural network output, a is the bias parameter, b is the output parameter from the nth hidden node, t_n is the output from the nth hidden node, and $g'(\cdot)$ is considered as the output function. The arbitrary functions $f'(\cdot)$ and $g'(\cdot)$ could be any function; however, the hyperbolic tangent function $(e^{-l} + e^l)/(e^l - e^{-l})$, the logistic function $e^l/(1 + e^l)$, or the linear function is the most common.

The critical characteristics list is trained and tested using the AB-ANN algorithm as shown in Algorithm 1. Assume that the input dataset comprises n features (f1, f2, …, fn). The information gain is calculated for each feature, indicating how much data is there in that feature set. The features with information gain higher than 0 are then considered and added to the list of important features. The accuracy of the classifier is then calculated. Remove each feature from the list of essential elements one by one. Then, train and test the remaining features through the ANN classifier model. If removing this characteristic affects classifier accuracy, it is the most important feature, and it is thus included to the list of the most important features. If removing this feature improves classifier accuracy, it is no longer a necessary feature and will be removed. This approach is continued till all key features have been tested and a list of the most important features has been created. The prediction is made as to whether or not the character would survive based on the most essential features from the feature list. This method can be used to save the lives of patients who have undergone transplant surgery.

4. Results and Discussions

This section evaluates the proposed AB-ANN method's performance. Two important indicators are used to evaluate performance in the test: the number of selected characteristics and predictive accuracy.

4.1. Performance Metrics. The following performance measures were used: true positive (tp): the model's predicted number of graft survival matches with the historical data; true negative (tn): the number of graft failures predicted by

the model matches with the historical data; false positive (fp): the number of grafts that the model predicts will survive although the prior examples have resulted in graft failure; false negative (fn): the number of graft failures predicted by the model when historical data have shown graft survival. After the computation of those metrics, the following measures are calculated. They are classification accuracy, precision, recall, F-measure, root mean square error, and mean absolute error.

The root mean square error and mean absolute error comparison of the proposed and existing methods is described in Table 1, and its pictorial representation is mentioned in Figure 2. From the figure, it is clear that the proposed method has minimum error rate compared with the existing mechanisms. The proposed method has lower root mean square error of 15.4% and lower mean absolute error of 9.3%.

4.1.1. Accuracy. The simplest intuitive performance metric is accuracy, which is defined as the ratio of precisely predicted observations to all observations. The proportion of accurately categorized patterns to the total number of classified patterns is known as accuracy. It is calculated using (6) as follows:

$$\text{Accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{fp} + \text{tn} + \text{fn}}. \qquad (6)$$

Table 2 and Figure 3 compare the suggested method's accuracy to that of the most recent techniques. Table 2 shows that the proposed AB-ANN predictive approach for renal transplantation could improve the classification accuracy rate while reducing the feature selection difficulty.

4.1.2. Precision. Precision is measured by the amount of positive class predictions which belongs to the positive class [28–30]. Precision is characterized as the proportion of the rate of correctly classified events in all detected events. It is computed using the following:

$$\text{Precision} = \frac{\text{tp}}{\text{tn} + \text{fp}}. \qquad (7)$$

The precision comparison of the proposed and existing methods is described in Table 3, and its pictorial representation is mentioned in Figure 4. From the figure, it is clear that the proposed AB-ANN method has higher precision value (97.6%) compared with the existing mechanisms. This shows the outperformance of the proposed method over existing mechanisms.

4.1.3. Recall. Recall is described as the amount of positive class predictions that are made of all positive examples in the dataset [31–33]. The fraction of right events among all events is known as recall. It is calculated using the following:

$$\text{Recall} = \frac{\text{tp}}{\text{tp} + \text{fn}}. \qquad (8)$$

```
Input: Input Dataset
Output: Input point classified as survive or not
The training samples are represented in n dimensional vector space
Consider a test data point P
Data processing operation
    Data cleaning to remove post-operative data like name, ID, etc.
    Data censoring to ensure the graft survival status
f1, f2, . . .., fn represents the feature set for the selected point from the n-dimensional space
  for each fn do
    Calculate the information gain (IG)
  if IG(fn) > 0
    Add the fn to feature list FL
  for each fn in FL do
Select the most important features fm using the fitness of ABO mechanism (1)
Remove fm from FL
    Compute accuracy for the selected fm A(fm) and accuracy of the remaining features in FL
    if A(fm) < A
        Include fm to the most important feature list
From the fm the test point P is classified as survive or not through the ANN model
```

ALGORITHM 1: Proposed AB-ANN algorithm.

TABLE 1: Error rate comparison of existing and proposed methods.

Methods	Root mean square error (%)	Mean absolute error (%)
Naïve Bayesian	57.44	38.79
J48	52.23	36.83
Random forest	41.99	34.08
Proposed (AB-ANN)	15.4	9.3
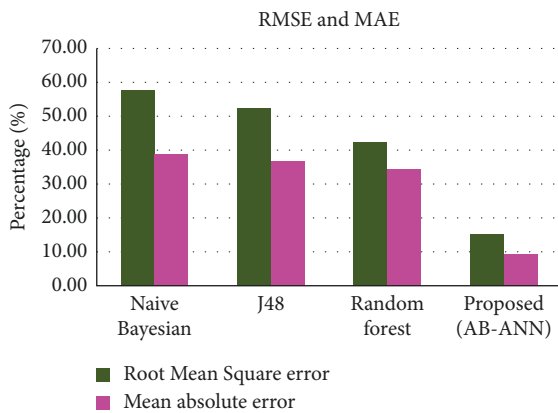
FIGURE 2: Comparison of RMSE and MAE.

TABLE 2: Accuracy comparison of existing and proposed methods.

References	Technique		Accuracy (%)
	Feature selection	Classification	
[19]	Data analytic method	Bayes net classifier	68.4
[27]	Kaplan–Meier	Nomogram	72
Proposed	IG + ABO	ANN	99.89
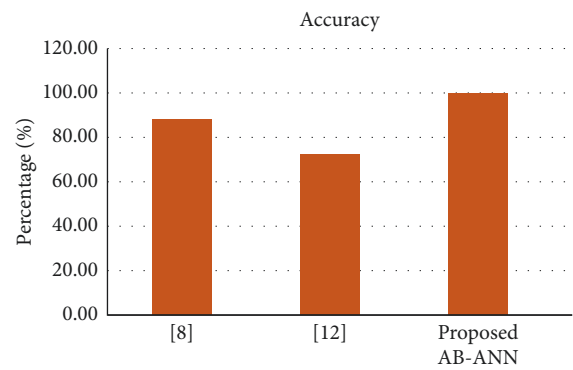
FIGURE 3: Comparison of accuracy.

TABLE 3: Precision comparison of existing and proposed methods.

Methods	Precision (%)
Naïve Bayesian	68.3
J48	63.1
Random forest	57.1
Proposed (AB-ANN)	97.6
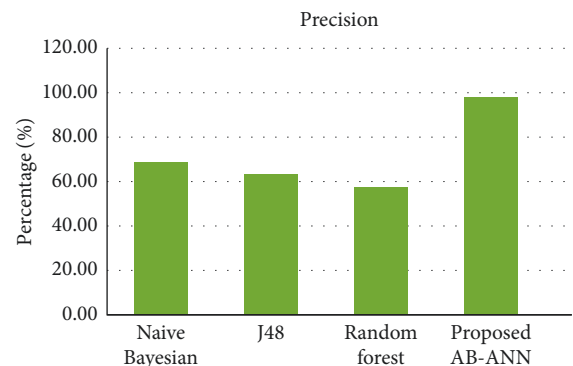
FIGURE 4: Comparison of precision.

TABLE 4: Recall comparison of existing and proposed methods.

Methods	Recall (%)
Naïve Bayesian	60.6
J48	68.1
Random forest	75.4
Proposed (AB-ANN)	98.2

The recall comparison of the proposed and existing methods is described in Table 4, and its pictorial representation is mentioned in Figure 5. From the figure, it is clear that the proposed AB-ANN method has higher recall value (98.2%) compared with other existing mechanisms.

4.1.4. F-Measure. It is the degree of harmonic mean among precision and recall. It is a statistical measure utilized to rate the performance. F1-score is formulated as follows:

$$F - \text{measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (9)$$

The F-measure comparison of the proposed and existing methods is described in Table 5, and its pictorial representation is mentioned in Figure 6. From the figure, it is clear that the proposed AB-ANN method has higher recall value (99.2%) compared with other existing mechanisms.

5. Discussion

The extensive availability of alternative treatments has increased the life span of patients having end-stage renal disease. The performance of the AB-ANN technique in detecting the survival rate of persons with kidney graft failure was compared with that of other methodologies in this research. The suggested prediction methodology is tested using the UNC dataset, and the results are compared with other recent methods. The predictions generated by ANN were more exact than previous techniques based on the evaluation parameters like accuracy, precision, recall, f-measure, and error rate.

Experiments demonstrated that the newly proposed kidney transplantation survival estimation technique surpassed all previous current strategies, with prediction accuracy and F-measure scores of 99.89 percent and 99.2 percent, respectively. The proposed prediction technique has achieved best accuracy, higher speed, and higher F-measure. Furthermore, the novel feature selection strategy has been successful in speeding up categorization by decreasing the amount of characteristics to a minimum. As a result, it is obvious that the proposed procedure is quite reliable and produces excellent outcomes. The nature of this model allows it to be utilized for both short and long-term forecasting.

Such predictive techniques could aid in the implementation of personalised treatment in kidney transplantation. It is stated that the innovative proposed prediction technique can increase classification accuracy while reducing feature selection complexity. These results show the efficacy of the proposed strategy. The proposed prediction
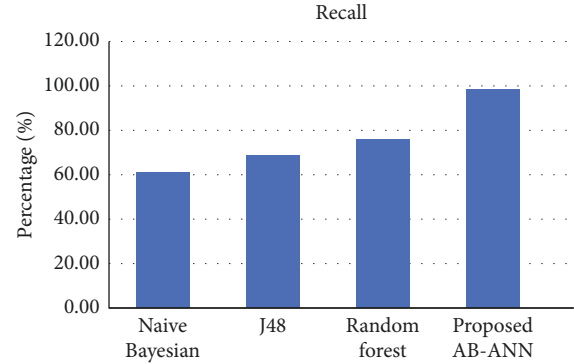
FIGURE 5: Comparison of recall.

TABLE 5: F-measure comparison of existing and proposed methods.

Methods	F-measure (%)
Naïve Bayesian	62.5
J48	64.3
Random forest	65
Proposed (AB-ANN)	99.2
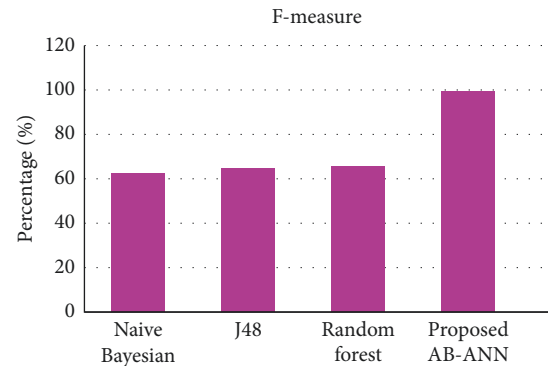
FIGURE 6: Comparison of F-measure.

model might be used to a variety of transplant datasets, according to the researchers.

6. Conclusion

The importance of predicting the outcome of kidney transplantation cannot be overstated. This will allow patients to choose the best accessible kidney donor and the best immunosuppressive medication. The ability to predict graft survival following transplanting is essential, and it is especially a challenging problem since it is important to understand the donor-recipient matching method. As finding donors is challenging, this matching is highly essential. Prediction of graft survival in kidney transplantation is a serious and therapeutically significant issue. An optimised deep learning framework for risk prediction of graft failure was built in this study, and it displayed a higher level of prediction performance. These algorithms outperformed those reported in the literature for existing risk prediction

tools, and the future research would focus on how to best integrate such models into healthcare algorithms to improve kidney recipients' long-term health.

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 to report regarding the present study.

Acknowledgments

References

[1] C. Legendre, G. Canaud, and F. Martinez, "Factors influencing long-term outcome after kidney transplantation," *Transplant International*, vol. 27, no. 1, pp. 19–27, 2014.

[2] M.-C. Fournier, Y. Foucher, P. Blanche et al., "Dynamic predictions of long-term kidney graft failure: an information tool promoting patient-centred care," *Nephrology Dialysis Transplantation*, vol. 34, no. 11, pp. 1961–1969, 2019.

[3] S. Kawakita, J. L. Beaumont, V. Jucaud, and M. J. Everly, "Personalized prediction of delayed graft function for recipients of deceased donor kidney transplants with machine learning," *Scientific Reports*, vol. 10, no. 1, Article ID 18409, 2020.

[4] R. Marcén, "Immunosuppressive drugs in kidney transplantation," *Drugs*, vol. 69, no. 16, pp. 2227–2243, 2009.

[5] S. Hoffmann, J. Park, L. M. Jacobson et al., "Donor genomics influence graft events: the effect of donor polymorphisms on acute rejection and chronic allograft nephropathy," *Kidney International*, vol. 66, no. 4, pp. 1686–1693, 2004.

[6] M. Coemans, C. Süsal, B. Döhler et al., "Analyses of the short- and long-term graft survival after kidney transplantation in Europe between 1986 and 2015," *Kidney International*, vol. 94, no. 5, pp. 964–973, 2018.

[7] D. A. Shoskes and J. M. Cecka, "Effect of Delayed Graft Function on Short-And Long-Term Kidney Graft Survival," *Clinical transplants*, vol. 31, pp. 297–303, 1997.

[8] P. Rashidi Khazaee, J. Bagherzadeh, Z. Niazkhani, and H. Pirnejad, "A dynamic model for predicting graft function in kidney recipients' upcoming follow up visits: a clinical application of artificial neural network," *International Journal of Medical Informatics*, vol. 119, pp. 125–133, 2018.

[9] M. R. First, "Renal function as a predictor of long-term graft survival in renal transplant patients," *Nephrology Dialysis Transplantation*, vol. 18, no. suppl_1, pp. i3–i6, 2003.

[10] B. Kaplan and J. Schold, "Neural networks for predicting graft survival," *Nature Reviews Nephrology*, vol. 5, no. 4, pp. 190–192, 2009.

[11] L. Tapak, O. Hamidi, P. Amini, and J. Poorolajal, *Prediction of Kidney Graft Rejection Using Artificial Neural Network*, vol. 23, no. 4, p. 8, 2017.

[12] M. A. Ghoneim, M. A. Bakr, A. F. Refaie et al., "Factors affecting graft survival among patients receiving kidneys from live donors: a single-center experience," *BioMed Research International*, vol. 2013, pp. 1–9, 2013.

[13] F. Poli, M. Scalamogna, M. Cardillo, E. Porta, and G. Sirchia, "An algorithm for cadaver kidney allocation based on a multivariate analysis of factors impacting on cadaver kidney graft survival and function," *Transplant International*, vol. 13, no. S1, pp. S259–S262, 2000.

[14] S. Krikov, A. Khan, B. C. Baird et al., "Predicting kidney transplant survival using tree-based modeling," *ASAIO Journal*, vol. 53, no. 5, pp. 592–600, 2007.

[15] V. Rao, R. S. Behara, and A. Agarwal, "Predictive modeling for organ transplantation outcomes," in *Proceedings of the 2014 IEEE International Conference on Bioinformatics and Bioengineering*, pp. 405–408, Boca Raton, FL, USA, November 2014.

[16] A. Oztekin, D. Delen, and Z. Kong, "Predicting the graft survival for heart-lung transplantation patients: an integrated data mining methodology," *International Journal of Medical Informatics*, vol. 78, no. 12, pp. e84–e96, 2009.

[17] M. E. Brier, P. C. Ray, and J. B. Klein, "Prediction of Delayed Renal Allograft Function Using an Artificial Neural Network," *Nephrol Dial Transplant*, vol. 18, p. 5, 2003.

[18] T. S. Brown, E. A. Elster, K. Stevens et al., "Bayesian modeling of pretransplant variables accurately predicts kidney graft survival," *American Journal of Nephrology*, vol. 36, no. 6, pp. 561–569, 2012.

[19] K. Topuz, F. D. Zengul, A. Dag, A. Almehmi, and M. B. Yildirim, "Predicting graft survival among kidney transplant recipients: a Bayesian decision support model," *Decision Support Systems*, vol. 106, pp. 97–109, 2018.

[20] K. D. Yoo, J. Noh, H. Lee et al., "A machine learning approach using survival statistics to predict graft survival in kidney transplant recipients: a multicenter cohort study," *Scientific Reports*, vol. 7, no. 1, pp. 8904–8912, 2017.

[21] R. S. Lin, S. D. Horn, J. F. Hurdle, and A. S. Goldfarb-Rumyantzev, "Single and multiple time-point prediction models in kidney transplant outcomes," *Journal of Biomedical Informatics*, vol. 41, no. 6, pp. 944–952, 2008.

[22] W. Duch, R. Adamczak, and K. Grabczewski, "A new methodology of extraction, optimization and application of crisp and fuzzy logical rules," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 277–306, 2001.

[23] S. A. A. Naqvi, K. Tennankore, A. Vinson, P. C. Roy, S. Sibte, and R. Abidi, "Predicting kidney graft survival using machine learning methods: prediction model development and feature significance analysis study," *Journal of Medical Internet Research*, vol. 23, p. 21, 2021.

[24] I. Eraky, H. A. El-Kappany, and M. A. Ghoneim, "Laparoscopic nephrectomy: Mansoura experience with 106 cases," *British Journal of Urology*, vol. 75, no. 3, pp. 271–275, 1995.

[25] B. Azhagusundari and A. S. Thanamani, "Feature selection based on information gain," *International Journal of Innovative Technology and Exploring Engineering*, vol. 2, no. 2, pp. 18–21, 2013.

[26] J. B. Odili, M. N. M. Kahar, and S. Anwar, "African buffalo optimization: a swarm-intelligence technique," *Procedia Computer Science*, vol. 76, pp. 443–448, 2015.

[27] A. Akl, A. Mostafa, and M. A. Ghoneim, "Nomogram that predicts graft survival probability following living-donor kidney transplant," *Experimental and Clinical Transplantation: Official Journal of the Middle East Society for Organ Transplantation*, vol. 6, no. 1, pp. 30–36, 2008.

[28] A. I. Taloba and S. S. I. Ismail, "An intelligent hybrid technique of decision tree and genetic algorithm for E-mail spam detection," in *Proceedings of the Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 99–104, Cairo, Egypt, December 2019.

[29] A. I. Taloba, M. R. Riad, and T. H. A. Soliman, "Developing an efficient spectral clustering algorithm on large scale graphs in spark," in *Proceedings of the 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 292–298, Cairo, Egypt, December 2017.

[30] M. H. R. Khalaf, Z. M. Abdel Azim, W. H. A. H. Elkhateeb, O. R. Shahin, and A. I. Taloba, "Explore the E-learning management system lower usage during COVID-19 pandemic," *Information Sciences Letters*, vol. 11, no. 2, pp. 537–548, 2022.

[31] A. El-Komy, O. R. Shahin, R. M. Abd El-Aziz, and A. I. Taloba, "Integration of computer vision and natural language processing in multimedia robotics application," *Information Sciences Letters*, vol. 11, no. 3, pp. 765–775, 2022.

[32] A. I. Taloba, A. A. Sewisy, and Y. A. Dawood, "Accuracy enhancement scaling factor of Viola-Jones using genetic algorithms," in *Proceedings of the 14th International Computer Engineering Conference (ICENCO)*, pp. 209–212, Cairo, Egypt, December 2018.

[33] M. M. Abdelgwad, T. H. A Soliman, A. I. Taloba, and M. F. Farghaly, "Arabic aspect based sentiment analysis using bidirectional GRU based models," *Journal of King Saud University - Computer and Information Sciences*, 2021.

Hindawi

Research Article

Data Confidentiality in Healthcare Monitoring Systems Based on Image Steganography to Improve the Exchange of Patient Information Using the Internet of Things

Hussah N. AlEisa

Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh 11671, Saudi Arabia

Correspondence should be addressed to Hussah N. AlEisa; haleisa@pnu.edu.sa

Recently, with the availability of fast and reliable Internet, the distance between a patient and a doctor is becoming unimportant. Physicians will be able to request the medical images of their patients regardless of the geographical area. However, a lot of challenges face such successful implementation. To facilitate remote diagnosis, patient electronic medical record (EMR), including medical images, that originates in one system needs to be exchanged either within the same organization or across different organizations. Steganography is the practice of concealing a secret message inside a cover medium. In this paper, steganography will be used to embed the patient's personal information securely and imperceptibly in their medical images to enhance confidentiality in case of a distant diagnosis. The security of the medical data is improved to maintain confidentiality and integrity using IoT. The least significant bit of the approximate coefficient of integer wavelet transform is proposed. The distortion between the cover image and stego-image is obtained by measuring the mean square error and PSNR, and normalized cross-correlation is utilized to estimate the degree of closeness between the cover image and stego-image.

1. Introduction

Communication through digitized media has been increasingly evident with the development of the Internet. All individual and commercial communication takes place on the Internet, where computerized media is the primary means. When sensitive data from businesses and organizations is shared, the security of the information is a major problem. They are required to keep their data safe from interfering eyes. However, everyone in today's world has access to the Internet, so there is a great danger in transmitting data digitally. The conservation of data during transmission is addressed in this way [1].

In today's world, the confidentiality of secret data is paramount, and advances in the security of computers have positioned steganography as a superior technique for acquiring secured data. Steganography is the method of concealing secret data in a message, audio file, picture, or video by embedding it in another image, audio file, video, or message [2–4]. It is used to keep sensitive information safe from hackers. Nowadays, the volume of data shared via the Internet is expanding. As a result, data security is considered a severe concern when data is communicated through the Internet [5, 6].

In steganography methods, each pixel of the cover picture is hidden with an equivalent number of secret bits. The embedding alteration in the cover image is equal. Individual pixels in a digital picture, on the other hand, have complicated statistical connections. As a result, the picture quality is automatically lowered while modifications with equal number of bits are made in the cover image pixel [7]. Different adaptive embedding techniques have been included within the steganography method to address these concerns. Each pixel of the cover picture is embedded with a changing number of bits using this adaptive embedding approach. As a result, the majority of researchers concentrated on adaptive strategies to increase the safety of the steganography approach [8–12]. Additionally, the value of

edge pixels is unaffected by modifications made during the embedding process. As a result, edge pixels can hold more hidden bits than smooth pixels. In most applications, metaheuristic algorithms are employed to tackle optimization problems [13–16].

To safeguard data, IoT transmitted data in the cloud through the Internet is employed. High data security is provided via complicated encryption and decryption technology. Only encryption and decryption are used, but data concealment offers a greater benefit [17, 18].

For security difficulties in the data communication between two devices in an IoT network as shown in Figure 1, several security criteria such as authentication, integrity, and secrecy were applied. Attacks are classified as low-, medium-, high-, and extremely high-level attacks based on their behavior and threat level.

The two types of image steganographic methods are spatial domain and frequency domain [19]. Spatial domain approaches deal with the direct change of picture pixels, and while they have a higher payload and imperceptibility, they are vulnerable to statistical assaults [20]. Frequency domain approaches, on the other hand, use changed coefficients resulting from different transformations such as DWT, DFT, and DCT for data embedding. These approaches are more resistant to image processing assaults, but they are computationally demanding and have a small payload, making them unsuitable for real-time applications [21, 22].

There are two ways to classify steganography techniques. Steganography methods are classified into image steganography, video steganography, text steganography, audio steganography, and network steganography, depending on the kind of cover image [23]. Steganography techniques are divided into two categories based on the embedding domains: spatial domain techniques, such as least significant bit and pixel value differencing approaches; transform domain techniques, such as discrete cosine transform (DCT), discrete wavelet transform (DWT), and integer wavelet transform (IWT); spread spectrum systems; masking distortion systems; and filtering methods [24].

2. Related Work

For securing information in an IoT architecture, three-color picture steganography algorithms are proposed. The first and third techniques employ red, green, and blue components for information transmission, whereas the second technique utilizes green and blue components. The dynamic positioning techniques were developed by utilizing the shared secret key to hide data in the deeper layer of the image channels [25].

The performance analysis of the secret image steganography technique for the security of images and data is discussed. For image steganography, modified LSB replacement and data mapping algorithms have been developed. Initially, the secret picture was preprocessed utilizing the data mapping method which was used to embed the secret picture in the cover image. In general, the majority of LSB approaches did not rely on pixel correlation or picture content. As a result, it might be detected through RS analysis.
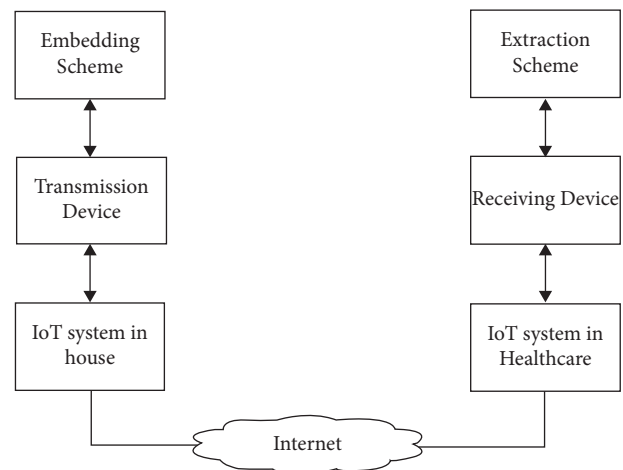
Figure 1: Data hiding and extraction for broadcasting between IoT devices.

Consequently, some preventive measures are required to enhance the security of LSB-based steganography approaches. An edge detection procedure has been implemented in the LSB-based steganography approach to address this problem. Furthermore, before incorporating the secret images into the cover image, they should be encrypted to increase security. Due to the lack of encryption methods preceding the embedding process, the security level of several current approaches has been decreased [26].

The data secured in the fog cloud IoT system is presented. The user in one region of the system uploads substantial data using the suggested quantum steganography convention and then transmits the protected data to the fog cloud. The intended receiver receives the data from the mist cloud and extracts the anticipated material using the specified extraction method [27].

A method for encrypting any form of an image, particularly medical images, has been developed. They intended to secure the integrity of electronic medical data while also sustaining its availability and authentication to ensure that only authorized personnel had access to it. In the first stage, the AES encryption technology was used. With seven values retrieved from the ear image as feature vectors, the ear print is also included. By delivering medical pictures over the Internet, the suggested approach increased the security of these images and protected them from unauthorized access [28].

The decision tree approach is used to present a novel method for shifting the medical data of the patient by securing the data into the medical cover image. The encryption is performed in the form of several chunks which are disseminated uniformly. Secret numbers are allocated to the cover image in the mapping method based on the breadth-first search to insert the data. Before embeddings, the data was encrypted using the RSA technique [29].

A safe approach that can meet the security and protection requirements while also overcoming the SPECS flaws is offered. In addition, this demonstrates the plan's success through execution assessments in terms of confirmation postponement and transmission overhead [30].

Developments and security adjustments attempting to solve the vulnerabilities of the security approach are presented to overcome the instinctive safety flaws of the 2-factor substantiation method. The proposed security enhancements may be used with the 2-factor authentication approach to achieve a more secure and robust two-factor client verification through WSNs [31]. For embedding reasons, the interblock approach is utilized. Images in the JPEG format are referred to as host or stego-images. This approach is only used to conceal patient information in medical JPEG photos. The difference in the coefficients is calculated using discrete cosine transform (DCT) [32].

In today's healthcare systems, Internet of Things (IoT) devices play a critical role. To incorporate patient data in any cover media, 2D DWT is applied. For the cover photos, grayscale and color photographs are utilized. Text data is encrypted using standard methods before being embedded in the cover medium. Various numerical methods are used to validate the imperceptibility of the cover image [33, 34].

On the segmented image, the region of the object and the reversible watermarking technique are employed. If image modes such as X-rays, magnetic resonance imaging (MRI), or computed tomography (CT) images have been tampered with or forged, the presented methodologies perform effectively to identify the tampering using the hash code. As medical systems are more prone to fabrication or manipulation, reversible watermarking methods are particularly useful [35, 36].

Steganography is a method for embedding data in many images, as compared to classical steganography, which uses just one image at a time for embedding. In the event of an exceptional state in the communication media during data transmission, secret data bits can be recovered from many shares. Compressed JPEG images are extensively useful for communication channels. An intermediate image is constructed before transmitting it to the channel, which is near enough to the stego-image [37].

A novel method for securing secret data in a fingerprint image created from a hidden message is proposed. Unlike traditional steganography techniques, there is no requirement for a cover signal for the embedding process. The secret message is transferred to the polynomial, encoded at diverse points of polarities, and utilized as a portion of the hologram to generate the fingerprint image [38–40].

3. Methodology

This section discusses the proposed methodology for maintaining data confidentiality during the IoT distribution process. Because information is passed across numerous hops in the Internet of Things, data security is critical. Due to the ease with which data may be accessed, the mixture of various gadgets and the interconnections established through a multitude of data give space for privacy breaches in IoT. As a result, in such a case, data may be secured by a reliable encoding technique. Accordingly, this study proposes a dependable data transmission model for a safe IoT connection, as shown in Figure 2, which is a depiction of a setup in healthcare that uses an IoT dispersed structure.

3.1. Steganography. A steganography approach based on encryption is proposed for conveying secret data. Normally, a digital image consists of disparate picture parts known as pixels. As a cover picture, a grayscale image and a color image are utilized in this work. As a result, a picture is represented by a large array of bytes. Image encryption, embedding phase, quality improvement, and extraction phase are the four key aspects of the system proposed. This programme is commonly used in photos, although the method's characteristics are generally stated in some figures, including hash marking. Steganography protects against unauthorized users and illegal copyrights.

Steganography is a progression in which secret data is hidden so that its presence cannot be recognized. This is why steganography is sometimes referred to as "covered writing." The goal of steganography is not only to secure the encryption but also to hide it so that no one can detect or determine the presence of the hidden secret data. This system or technique aims to hide the presence of any secret data. The person who is not permitted to push for knowledge access should not even know if any secret information is available.

The basic components of steganography are the message, the carrier and the stego-key. The message is that the secret text, image, video, or audio has to be safeguarded using the steganography process. The carrier is the path or medium through which the key and, hence, the covered message are sent. The stego-key is the password by which confidential data is protected and exposed as shown in Figure 3.

3.2. Image Encryption. In the encryption process, the hidden image is processed with binary by plane decomposition which is utilized to decompose the image into binary bit planes. The image is represented with binary planes in this method for a decimal number which is given as

$$
\begin{aligned}
B &= \sum_{i=0}^{n-1} b_i 2^i, \\
B &= b_0 2^0 + b_1 2^1 + \cdots\cdots\cdots + b_{n-1} 2^{n-1}, \\
B &= b_0 + 2b_1 \cdots\cdots\cdots\cdots + 2^{n-1} b_{n-1}.
\end{aligned}
\tag{1}
$$

The grayscale image has the pixel value in the range of 0 to 255, which is decomposed into binary bit values. With the support of secret key binary, keystreams are generated. These binary keystreams enter the two stages of the encryption model. The piecewise linear chaotic map is utilized to produce the keystreams, which are represented as

$$
x_{i+1} = \begin{cases} \dfrac{x_i}{\delta}, & 0 \le x_i \le \delta, \\[2mm] \dfrac{x_i - \delta}{0.5 - \delta}, & \delta \le x_i \le 0.5, \\[2mm] F(1 - x_i, \delta), & 0.5 \le \delta \le 1. \end{cases}
\tag{2}
$$
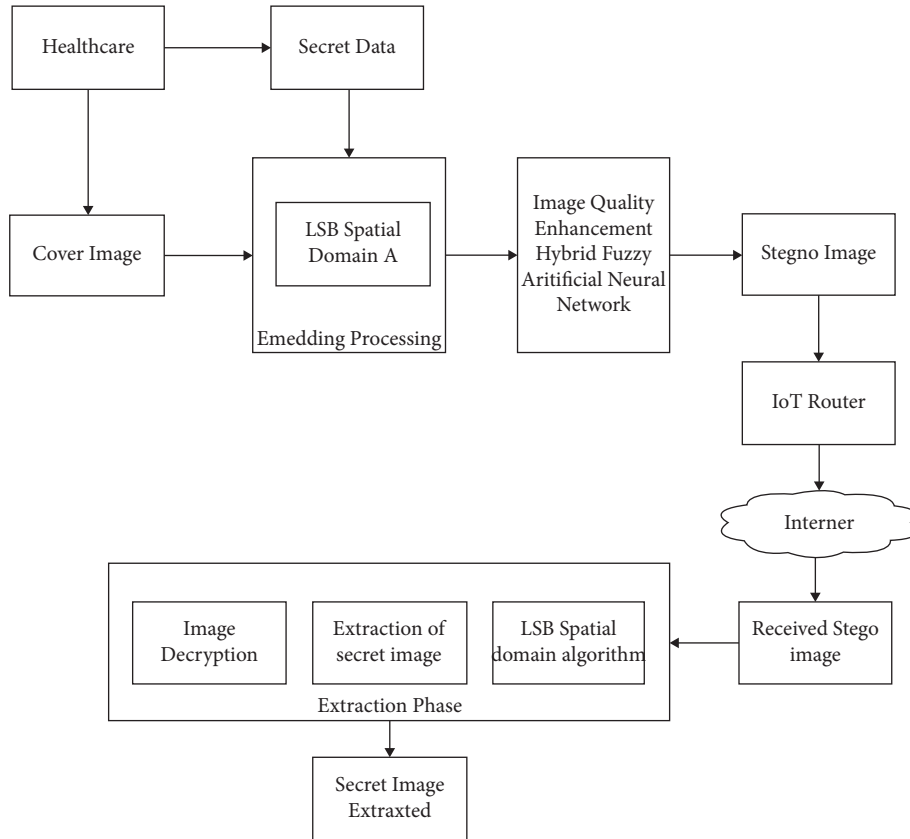
FIGURE 2: Block diagram of the proposed methodology.
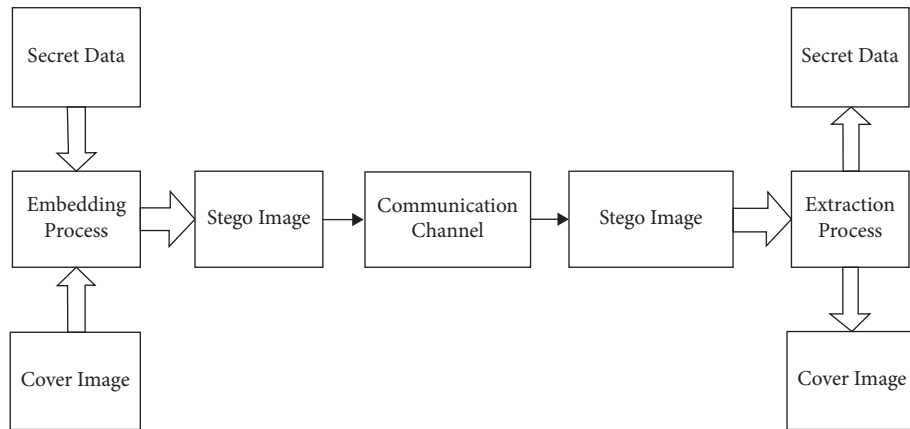
FIGURE 3: Steganography scheme.

The control parameter is represented as δ, and x_i provides the initial condition of piecewise linear chaotic mapping. The set of the hidden image δ is encrypted with the initial value x_0

The keystream $X = \{x_1, x_2, \ldots \ldots x_n\}$ is converted into the integer sequence $X_1(i)$.

$$X_1 = \mathrm{mod}\left(\mathrm{floor}\left(X \times 10^{14}\right), 256\right). \qquad (3)$$

The bit-plane decomposition utilizes the keystream into bit planes to obtain the binary sequence k_1 and k_2 in which the bits are arranged from a higher bit plane to a lower bit plane.

3.2.1. Diffusion Stage. The diffusion stage is performed by the following steps.

(1) Elements M are added as

$$S_1 = \sum_{i-0}^{n-1} M_1(i). \qquad (4)$$

(2) Cyclic operation is performed in M_1 to obtain the matrix M_{11}, and M_1 element is shifted right by S_1 bits.

(3) First element M_1 is encrypted, and the key of the first element M_1 is given as

$$P_1(i) = M_{11}(i) \oplus M_{11}(i-1) \oplus M_2(i) \oplus k_1(i). \quad (5)$$

(4) Element P is given as

$$S_2 = \sum_{i=1}^{n-1} P_1(i). \quad (6)$$

(5) Cyclic operation is performed in M_2 to obtain the matrix M_{22}, and M_2 element is shifted right by S_2 bits.

(6) First element M_{22} is encrypted, and the key of the first element M_2 is given as

$$P_2(i) = M_{22}(i) \oplus M_{22}(i-1) \oplus P_1(i) \oplus k_1(i). \quad (7)$$

3.2.2. Confusion Stage.
The steps performed in confusion matrix are given as follows.

(1) The elements P_1 and P_1 are added as given below:

$$S_3 = \sum_{i=0}^{nm} P_1(i) + P_2(i). \quad (8)$$

(2) The keystreams k_1 and k_2 are generated using the secret key $k(x, \delta)$. The initial value a_0 is generated using the following:

$$a_0 = \text{mod}\left(\frac{a_0 + S_3}{nm, 1}\right). \quad (9)$$

(3) The chaotic sequence is generated as

$$
\begin{aligned}
A_1 &= \{a_1, a_2, \ldots \ldots \ldots a_{nm}\}, \\
A_2 &= \{a_{nm+1}, a_{nm+2}, \ldots \ldots \ldots a_{nm}\}.
\end{aligned} \quad (10)
$$

(4) The integer sequence X_1 and X_2 is given as

$$
\begin{aligned}
X_1 &= \text{mod}\left(\text{floor}\left(A_1 \times 10^{14}\right), 4nm\right) + 1, \\
X_2 &= \text{mod}\left(\text{floor}\left(A_2 \times 10^{14}\right), 4nm\right) + 1.
\end{aligned} \quad (11)
$$

(5) The row vector R_1 is obtained by encrypting the swapping elements P_1 and P_2:

$$
\begin{aligned}
\text{temp} &= P_1(i), \\
P_1(i) &= P_2(X_1(i)), \\
P_2(X_1(i)) &= \text{temp}.
\end{aligned} \quad (12)
$$

(6) The row vector R_2 is obtained by encrypting the swapping elements P_1 and P_2:

$$
\begin{aligned}
\text{temp} &= P_2(i), \\
P_2(i) &= P_1(X_2(i)), \\
P_1(X_2(i)) &= \text{temp}.
\end{aligned} \quad (13)
$$

(7) R_1 and R_2 are the row vectors which are transformed into $n \times m$ images to obtain the secret image.

3.3. Embedding Process.
The embedding approach involves some cover image and secret image preparation, as well as secret key extraction and data hiding. The cover picture was chosen based on certain conclusions drawn from earlier steganography studies. This should be done carefully so that the superiority of the stego-image created after hiding is preserved. Certain pixels or blocks are chosen from the cover image using a random key. Before embedding, the secret picture is compressed and encrypted. Compression reduces the quantity of data to be hidden, while encryption improves security. Even if the primary concern of steganography is exploited, data should not be exposed. The secret image is compressed using a sophisticated wavelet-based compression algorithm. Simple bit operations like AND and OR are used to encrypt data. After that, the secret picture is transformed into a bitstream, referred to as the secret data. The LSB approach is utilized to disguise the secret data in the chosen pixels. It is sufficient to swap the final two bits if the quantity of data is less. Otherwise, the secret data is swapped for the least significant 3 bits of the chosen pixels to generate the stego-image as shown in Figure 4.

3.4. LSB Domain Algorithm.
The algorithm for hiding a hidden text in an image is the LSB. The LSB embedding technique uses the secret text bitstream to be hidden to substitute the LSBs of the pixels in the cover picture. Because deviations in the LSBs of pixels do not produce variation in the image, the stego-image is virtually identical to the cover image.

The pixel value $I(a, b)$ of LSB is similar to message bit which is embedded in $I(a, b)$, and it remains unchanged. The stego-image is obtained as follows:

$$
I_s(a, b) = \begin{cases}
I(a, b) - 1, & m = 0, \text{LSB}(I(a, b)) = 1, \\
I(a, b), & \text{LSB}(I(a, b)) = m, \\
I(a, b) + 1, & m = 0, \text{LSB}(I(a, b)) = 1,
\end{cases} \quad (14)
$$

where m is the next bit for embedding each pixel by changing a bit.

The pixels of an image must be adjusted to incorporate a hidden message. It is hard to differentiate between the cover image and the stego-image. This approach often generates significant distortion in the cover image when the number of hidden bits for each pixel reaches three. There are many steganographic to be utilized to mitigate the distortion induced by LSB replacement. Adaptive approaches alter the number of concealed bits in each pixel, resulting in a higher image quality compared to systems that rely only on LSB replacement. However, this comes at the expense of lowering the embedding capacity.

3.5. IWT Technique.
An integer data set is transformed into another integer data set using the IWT. When the data is hidden in the coefficients of the wavelet filters used in the
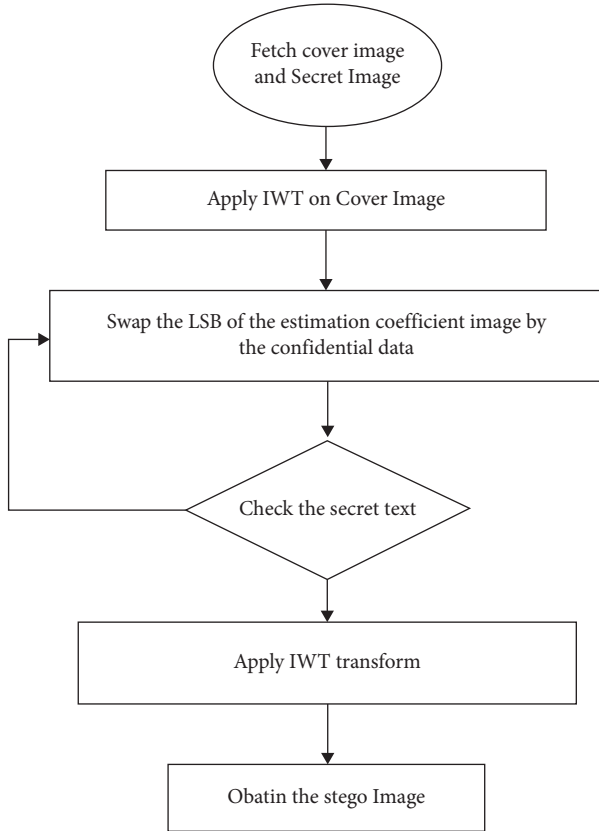
Figure 4: Flowchart of the LSB-IWT embedding process.

DWT, any method that is most effective for the floating-point values of the pixels that should be integers may result in the loss of the hidden data, failing the data hiding system. To avoid difficulties with wavelet filters' floating-point accuracy when the input data is an integer, such as in digital images, the output data will no longer be an integer, preventing perfect recreation of the original image and preventing information loss through forward and inverse transforms. The lifting technique is one of the approaches for performing the IWT. IWT can transform integer wavelet coefficients from pixel values and recreate the image from integer wavelet coefficients due to its numerical advantages.

The image pixel is decomposed into four subband wavelets of DWT: LL, LH, HL, and HH. The coefficient of the image is approximated using the LL subband, vertical details of the image are given using the LH subband, horizontal details of the image are examined in HL, and diagonal details of the image are given using the HH subband. The coefficient of IWT is computed as

$$LL_{a,b} = \left[\frac{\left(I_{2a,2b} + I_{2a,2b}\right)}{2} \right],$$

$$LH_{a,b} = I_{2a,2b+1} - I_{2a,2b},$$

$$HL_{a,b} = I_{2a+1,2b} - I_{2a,2b}, \tag{15}$$

$$HH_{a,b} = I_{2a+1,2b+1} - I_{2a,2b}.$$

The coefficient of inverse IWT is given as

$$I_{2a,2b} = LL_{a,b} - \left[\frac{HL_{a,b}}{2} \right],$$

$$I_{2a,2b+1} = LL_{a,b} - \left[\frac{HL_{a,b+1}}{2} \right], \tag{16}$$

$$I_{2a+1,2b} = I_{2a,2b+1} + LH_{a,b} - HL_{a,b},$$

$$I_{2a+1,2b+1} = I_{2a+1,2b} + HH_{a,b} + LH_{a,b}.$$

The steps of the algorithm for the embedding process in IWT-LSB are as follows:

(1) The cover image is read.

(2) The secret image is read.

(3) IWT is applied for the cover image.

(4) Change the LSB of the coefficient image by the secret image.

(5) Until the secret data is completely hidden in the cover image, step 4 will be continued.

(6) Inverse IWT is applied.

(7) Stego-image is obtained.

3.6. Image Quality Enhancement. The obtained stego-image from the embedding process is of insufficient quality. As a result, a processing procedure on the unique intelligent system is required. This phase is necessary for reducing the chances of numerical identification and other types of image modification attempts.

3.7. Hybrid Fuzzy Neural Network. An HFNN with a backpropagation learning method is employed to improve the image quality in this study. In general, neural networks resemble HFNNs that are inhomogeneous. The neural network refers to a framework that can simulate how the human brain learns. The stego-image transformed to binary bit values in order to identify free bits and bits that encompass secret bits. A buffer is built to keep the free bits that are not used by the embedding process. The stego-and cover images are also used to extract statistical and perceptual attributes. The statistical and visual aspects are represented by the chi-square probability and the Euclidian norm. The HFNN is provided with the free bits buffer as well as the two characteristics of the stego-image. The cover attributes are then compared to the HFNN outputs. By integrating the changed free bits with the secret bits, a new stego-image is created if the outputs match the characteristics of the cover image. The HFNN weights are modified using a backpropagation learning process.

(1) The stego-image is generated by hiding the secret image, which is implemented using embedding approach algorithms.

(2) Features are extracted from the stego-image using the feature extraction technique.

(3) A buffer is generated which is not utilized in the steganographic algorithm. The secret is not hidden in the buffer bit, and it is called free bits.

(4) The statistical and visual measure of the stego-image is measured using a fuzzy neural network with backpropagation. The statistical and visual measure of the stego-image is measured using a fuzzy neural network with backpropagation. Therefore, for an updated stego-image, the free bit buffer, statistical and visual measures are contained in the output layer.

(5) The output of the fuzzy network is compared with the cover image. The stego-image is formed if the output and cover image get matched and the output with the free bit is used by assembling the other bit in which the secret image is hidden. Otherwise, step 4 is repeated.

The input layer, rule layer, fuzzification layer, inference layer, and defuzzification layer are the five layers that compose HFNN as shown in Figure 5. The input neurons of the HFNN were trained using five layers of backpropagation, using inputs from the free bits buffer, and numerical and graphic characteristics. Following that, all hidden neurons in the fuzzification layer get the inputs. The membership function is used to perform fuzzy process on input characteristics at this layer. For the excellent approximation of input space, the Gaussian membership function is utilized. By altering the parameter values, this bell-shaped function produces several membership functions for the input characteristics. The strength of fuzzy rules is determined in the rule layer using the logical AND operator. On fuzzy inference, the inference layer executes OR operations. The HFNN output will emerge from the defuzzification layer. The weight parameter is used to link the nodes of all layers in HFNN.

The dimensionality reduction is a feature extraction in which set of features is transformed form stego image. The statistical features are obtained from chi-square probability and visual features from the Euclidean norm. Let x and y be the input, then linguistic input variable A_1, A_1 and B_1, B_1. The linguist state is given as

$$A_j(u) = \exp\left[-\frac{1}{2}\left(\frac{a - u_{j1}}{v_{j1}}\right)^2\right],$$

$$B_j(u) = \exp\left[-\frac{1}{2}\left(\frac{a - u_{j2}}{v_{j2}}\right)^2\right]. \tag{17}$$

$\{u_{j1}, u_{j1}, u_{j1}, u_{j1}\}$ are the parameter set. The logical operator is used to strength the output of the network layer:

$$F_1 = A_1(x_0) \Lambda B_1(y_0),$$
$$F_2 = A_2(x_0) \Lambda B_2(y_0). \tag{18}$$

The defuzzification is obtained as the normalization:

$$F_{d1} = \frac{F_1}{F_1 + F_2},$$

$$F_{d2} = \frac{F_2}{F_1 + F_2}. \tag{19}$$

The error function is given by

$$E = \frac{1}{2}(y - 0)^2. \tag{20}$$

The desired output is represented as y.

3.8. Extraction Process.

The extraction process is used to extract the hidden image from the embedded process in adjustable order. The cover image used in the first step is not used to extract the secret image. The data is provided by the LSB, and the procedure was enhanced using the secret key.

3.8.1. Algorithm for Extraction Process in IWT-LSB

(1) Stego-image is read.

(2) Median filter is applied.

(3) IWT is applied for stego-image.

(4) The secret data is extracted for the approximate image coefficient of stego-image.

(5) Until the secret data is extracted, step 4 will be continued.

(6) Inverse IWT is applied.

(7) Image gets extracted.

(8) Extract the secret data.

3.9. Special Cases

3.9.1. IWT-LSB Algorithm for Grayscale Image.

The proposed hybrid IWT-LSB technique for grayscale images involves an embedding phase and an extraction phase as shown in Figure 6. In the embedding phase, the grayscale cover image is transformed using IWT, then the secret text is embedded in the LSBs of the cover image's coefficients, and finally the inverse IWT is used to construct the stego-image. Without knowing anything about the original image, the hidden text might be extracted throughout the extraction process. The hidden secret text is retrieved from the LSBs of the filtered stego-image's coefficients, and the inverse IWT is useful for creating the extracted image.

The 512×512 bitmap grayscale images are utilized as cover images for the hybrid IWT-LSB algorithms on grayscale images. The size of the estimated coefficient images after performing the IWT is 256×256, which implies that a secret text with up to 8 and 192 digits may be secured. To improve the system's robustness and secure the message from external impacts such as noise, compression, and filtering, the secret data is placed in the LSBs of the estimated coefficient images of the cover images.

3.9.2. IWT- LSB Algorithm for Color Image.

The color image is split into R, G, and B components in the proposed methodology, and the three components are employed to hide data as shown in Figure 7. In the embedding phase, the RBG component is used to decompose the color of cover image in which the IWT transform is used with the signature of the user and the secret data is embedded with actual
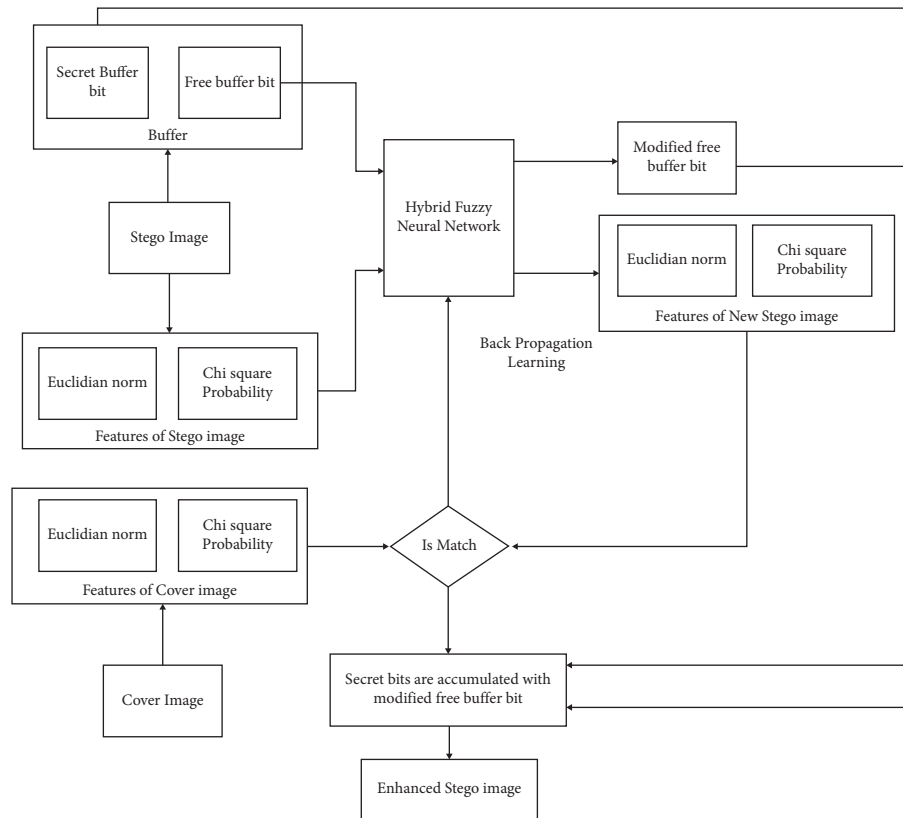
FIGURE 5: Quality enhancement of stego-image using HFNN.

(a)

(b)
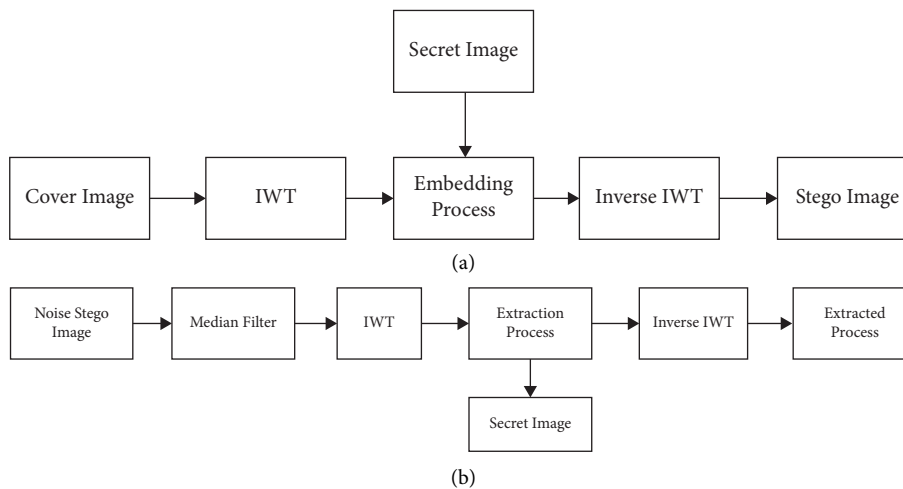
FIGURE 6: Proposed embedding and extraction process for grayscale image steganography.

length in LSB and coefficient image is approximated for the red component of cover image. The image's approximated green and blue components coefficient are used to secure the secret data. Then, the inverse IWT is applied to each component once the embedding process is complete, and then these components are recombined to generate a stego-image.

The hidden data may be extracted during the extraction process without knowing anything about the original image. The R, G, and B components are decomposed from the noisy stego-image in which median filter is utilized for filtering

and then transformed using IWT. The LSB of the approximated image coefficient is used for extracting the actual length of the secret image by utilizing the G and B components, and the inverse IWT is utilized to extract the original image.

4. Result and Discussion

The result of the IWT-LSB algorithm for the steganography image such as grayscale image and color image is analyzed.

(a)

(b)
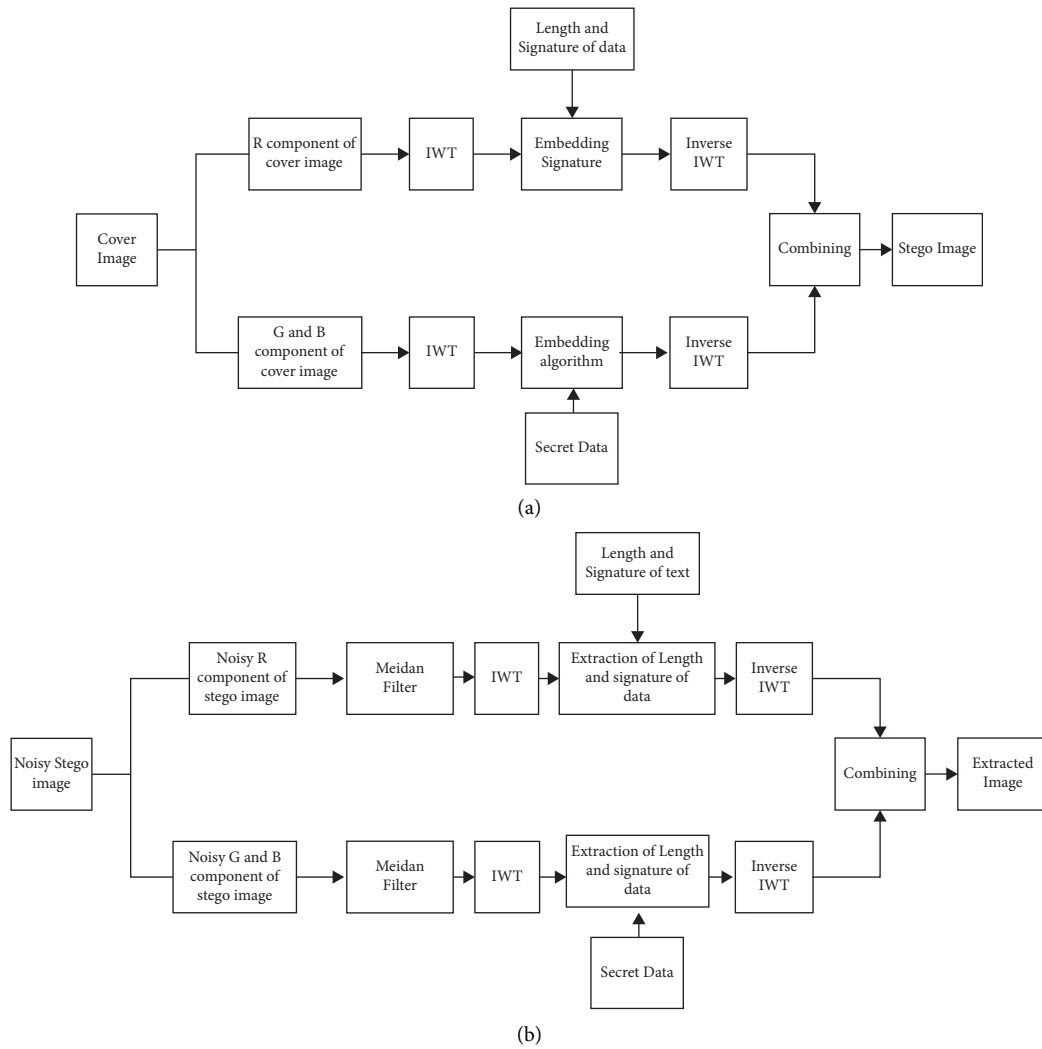
FIGURE 7: Proposed embedding and extraction process for color image steganography.

Parameters such as MSE, NCC, and PSNR are used for the performance evaluation. The error among the cover image and stego-image is examined using MES and PSNR and compared with the existing techniques.

The cover image is represented as I_C, and stego-image is represented as I_S. The mean square error is calculated using the following equation:

$$\text{MSE} = \frac{1}{nm} \sum_{a=0}^{m-1} \sum_{b}^{n-1} \left[I_S(a,b) - I_C(a,b) \right]^2. \tag{21}$$

The peak to signal noise ratio is given as

$$\text{PSNR} = 10 \, \log\left(\frac{I_{\max}^2}{\text{MSE}} \right). \tag{22}$$

The number of rows and columns is represented as n and m, and $I_{\max}$ is the maximum hold of the original image.

The similarity and dynamic extent of the cover image and secret image are quantified using the mean square error. The mean square error in the total number of pixels in color and grayscale image is given in Figure 8.
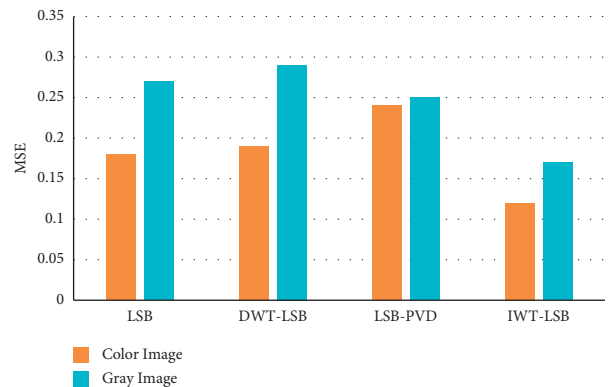
FIGURE 8: Comparison of mean square in color and grayscale image using proposed and existing technique.

The PSNR of the suggested algorithm compared with the existing technique is given in Figure 9. The PSNR regulates the difference in the dynamic range of invisibility in cover image and secret image in which the value is greater than
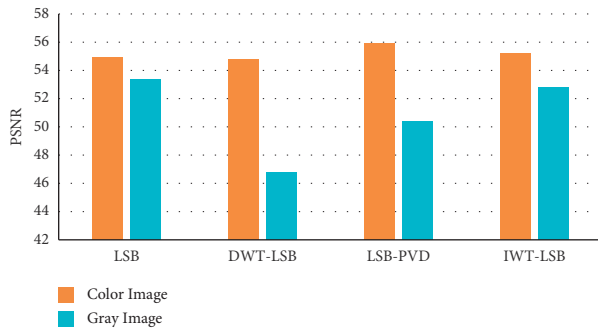
FIGURE 9: Comparison of PSNR in color and grayscale image using proposed and existing technique.
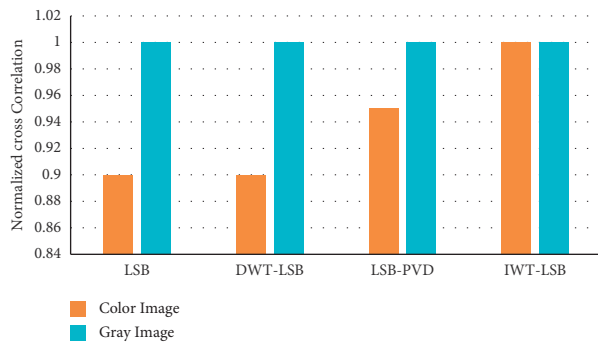
FIGURE 10: Comparison of NCC in color and grayscale image using proposed and existing technique.

53 dB in grayscale and color image compared to the existing techniques.

The degree of closeness between the cover image and stego-image is obtained using NCC which is shown in Figure 10. The degree of closeness is obtained after embedding the data in the secret image.

5. Conclusion

This paper used image steganography to securely and imperceptibly embed the patient's personal information in their medical images to enhance confidentiality in case of distant diagnosis. The least significant bit of the approximate coefficient of integer wavelet transform is proposed. This technique is analyzed for grayscale image and color image. IWT is utilized to hide the secret image in LSB in the grayscale image, while IWT with R, B, and G component is used for hiding the secret image in color image. The distortion between the cover image and stego-image is obtained by measuring the mean square error and PSNR, and the degree of closeness between the cover image and stego-image is estimated by utilizing the normalized cross-correlation. The result shows that the IWT-LSB technique can hide secret data with large length with better MSE, PSNR, and NCC.

Data Availability

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

Conflicts of Interest

The author declares that she has no conflicts of interest to report regarding the present study.

References

[1] A. A. J. Altaay, S. B. Sahib, and M. Zamani, "An introduction to image steganography techniques," in *Proceedings of the International Conference on Advanced Computer Science Applications and Technologies*, pp. 122–126, ACSAT, Kuala Lumpur, Malaysia, November 2012.

[2] K. A. Al-Afandy, O. S. Faragallah, A. ELmhalawy, E.-S. M. El-Rabaie, and G. M. El-Banby, "High security data hiding using image cropping and LSB least significant bit steganography," in *Proceedings of the 4th IEEE International Colloquium on Information Science and Technology (CiSt)*, pp. 400–404, IEEE, Tangier, Morocco, October 2016.

[3] A. Elhadad, S. Hamad, A. Khalifa, and A. Ghareeb, "High capacity information hiding for privacy protection in digital video files," *Neural Computing & Applications*, vol. 28, no. 1, pp. 91–95, 2017.

[4] S. Hamad, A. Khalifa, and A. Elhadad, "A blind high-capacity wavelet-based steganography technique for hiding images into other images," *Advances in Electrical and Computer Engineering*, vol. 14, no. 2, pp. 35–42, 2014.

[5] M. Hashim and M. Rahim, "Image steganography based on odd/even pixels distribution scheme and two parameters random function," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 22, 2017.

[6] B. M. Krishna, C. Santhosh, S. Suman, and S. K. Shireen, "Systems, and computers, "evolvable hardware-based data security system using image steganography through dynamic partial reconfiguration," *Journal of Circuits, Systems, and Computers*, vol. 31, no. 1, Article ID 2250014, 2022.

[7] C. Qin, W. Zhang, F. Cao, X. Zhang, and C.-C. Chang, "Separable reversible data hiding in encrypted images via adaptive embedding strategy with block selection," *Signal Processing*, vol. 153, pp. 109–122, 2018.

[8] Q. Li, X. Liao, G. Chen, and L. Ding, "A novel game-theoretic model for content-adaptive image steganography," in *Proceedings of the IEEE 37th International Conference on Distributed Computing Systems Workshops*, pp. 232–237, ICDCSW, Atlanta, GA, USA, June 2017.

[9] D. B. Khadse and G. Swain, "Data hiding using quotient value differencing and remainder value substitution avoiding incorrect extraction problem," *Sensing and Imaging*, vol. 22, no. 1, pp. 1–21, 2021.

[10] R. Sonar and G. Swain, "A hybrid steganography technique based on RR, AQVD, and QVC," *Information Security Journal: A Global Perspective*, pp. 1–20, 2022.

[11] R. Sonar and G. Swain, "Steganography based on quotient value differencing and pixel value correlation," *CAAI Transactions on Intelligence Technology*, vol. 6, no. 4, pp. 504–519, 2021.

[12] G. Swain and A. Pradhan, "Image steganography using remainder replacement, adaptive QVD and QVC," *Wireless Personal Communications*, vol. 123, no. 1, pp. 273–293, 2022.

[13] S. I. Nipanikar, V. Hima Deepthi, and N. Kulkarni, "A sparse representation based image steganography using particle swarm optimization and wavelet transform," *Alexandria Engineering Journal*, vol. 57, no. 4, pp. 2343–2356, 2018.

[14] G. Swain, "Adaptive and non-adaptive PVD steganography using overlapped pixel blocks," *Arabian Journal for Science and Engineering*, vol. 43, no. 12, pp. 7549–7562, 2018.

[15] G. Swain, "A data hiding technique by mixing MFPVD and LSB substitution in a pixel," *Information Technology and Control*, vol. 47, no. 4, pp. 714–727, 2018.

[16] G. Swain, "Two new steganography techniques based on quotient value differencing with addition-subtraction logic and PVD with modulus function," *Optik*, vol. 180, pp. 807–823, 2019.

[17] S. Arunkumar, V. Subramaniyaswamy, V. Vijayakumar, N. Chilamkurti, and R. Logesh, "SVD-based robust image steganographic scheme using RIWT and DCT for secure transmission of medical images," *Measurement*, vol. 139, pp. 426–437, 2019.

[18] E. S. B. Hureib and A. A. Gutub, "Enhancing medical data security via combining elliptic curve cryptography with 1-LSB and 2-LSB image steganography," *International J Comp Sci Network Security (IJCSNS)*, vol. 20, no. 12, pp. 232–241, 2020.

[19] A. Anees, A. M. Siddiqui, J. Ahmed, and I. Hussain, "A technique for digital steganography using chaotic maps," *Nonlinear Dynamics*, vol. 75, no. 4, pp. 807–816, 2014.

[20] Z. Xia, X. Wang, X. Sun, Q. Liu, and N. Xiong, "Steganalysis of LSB matching using differences between nonadjacent pixels," *Multimedia Tools and Applications*, vol. 75, no. 4, pp. 1947–1962, 2016.

[21] M. Sajjad, K. Muhammad, S. W. Baik et al., "Mobile-cloud assisted framework for selective encryption of medical images with steganography for resource-constrained devices," *Multimedia Tools and Applications*, vol. 76, no. 3, pp. 3519–3536, 2017.

[22] A. Elhadad, A. Ghareeb, and S. Abbas, "A blind and high-capacity data hiding of DICOM medical images based on fuzzification concepts," *Alexandria Engineering Journal*, vol. 60, no. 2, pp. 2471–2482, 2021.

[23] S. Kaur, S. Bansal, and R. K. Bansal, "Steganography and classification of image steganography techniques," in *Proceedings of the International Conference on Computing for Sustainable Global Development*, pp. 870–875, IEEE, New Delhi, India, March 2014.

[24] V. Thanikaiselvan and P. Arulmozhivarman, "High security image steganography using IWT and graph theory," in *Proceedings of the IEEE International Conference on Signal and Image Processing Applications*, pp. 337–342, IEEE, Melaka, Malaysia, October 2013.

[25] A. Abdelaziz, M. Elhoseny, A. S. Salama, and A. M. Riad, "A machine learning model for improving healthcare services on cloud computing environment," *Measurement*, vol. 119, pp. 117–128, 2018.

[26] A. Arya and S. Soni, "Performance evaluation of secrete image steganography techniques using least significant bit (LSB) method," *International Journal of Computer Science Trends and Technology*, vol. 6, no. 2, pp. 160–165, 2018.

[27] A. A. A. El-Latif, B. Abd-El-Atty, M. S. Hossain, S. Elmougy, and A. Ghoneim, "Secure quantum steganography protocol for fog cloud internet of things," *IEEE Access*, vol. 6, pp. 10332–10340, 2018.

[28] A. S. Anwar, K. K. A. Ghany, and H. E. Mahdy, "Improving the security of images transmission," *International Journal of Bio-Medical Informatics and E-Health*, vol. 3, no. 4, pp. 7–13, 2015.

[29] M. Jain, R. C. Choudhary, and A. Kumar, "Secure medical image steganography with RSA cryptography using decision tree," in *Proceedings of the 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, pp. 291–295, Noida, India, December 2016.

[30] S.-J. Horng, S.-F. Tzeng, Y. Pan et al., "b-SPECS+: batch verification for secure pseudonymous authentication in VANET," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 11, pp. 1860–1875, 2013.

[31] M. K. Khan and K. Alghathbar, "Cryptanalysis and security improvements of 'two-factor user authentication in wireless sensor networks'," *Sensors*, vol. 10, no. 3, pp. 2450–2459, 2010.

[32] X. Liao, J. Yin, S. Guo, X. Li, and A. K. Sangaiah, "Medical JPEG image steganography based on preserving inter-block dependencies," *Computers & Electrical Engineering*, vol. 67, pp. 320–329, 2018.

[33] M. Elhoseny, G. Ramirez-Gonzalez, O. M. Abu-Elnasr, S. A. Shawkat, N. Arunkumar, and A. Farouk, "Secure medical data transmission model for IoT-based healthcare systems," *IEEE Access*, vol. 6, pp. 20596–20608, 2018.

[34] H. E. Rostam, H. Motameni, and R. Enayatifar, "Privacy-preserving in the Internet of Things based on steganography and chaotic functions," *Optik*, vol. 258, Article ID 168864, 2022.

[35] R. Eswaraiah and E. Sreenivasa Reddy, "Robust medical image watermarking technique for accurate detection of tampers inside region of interest and recovering original region of interest," *IET Image Processing*, vol. 9, no. 8, pp. 615–625, 2015.

[36] N. Sahu, D. Peng, and H. Sharif, "Diagnosis-steganography-transmission: an innovative integrated paradigm for ECG healthcare," *SN Computer Science*, vol. 2, no. 4, pp. 1–22, 2021.

[37] S. Al-Refai and M. M. Al-Jarrah, "Secure data hiding technique using batch video steganography," in *Proceedings of the 2nd International Conference on Information Hiding and Image Processing*, pp. 1–4, ACM, London, United Kingdom, September 2019.

[38] S. Li and X. Zhang, "Toward construction-based data hiding: from secrets to fingerprint images," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1482–1497, 2018.

[39] A. I. Taloba, A. A. Sewisy, and Y. A. Dawood, "Accuracy enhancement scaling factor of Viola-Jones using genetic algorithms," in *Proceedings of the 14th International Computer Engineering Conference (ICENCO)*, pp. 209–212, Giza, Egypt, December 2018.

[40] A. I. Taloba and S. S. Ismail, "An intelligent hybrid technique of decision tree and genetic algorithm for e-mail spam detection," in *Proceedings of the 9th International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 99–104, Cairo, Egypt, December 2019.

Hindawi

Research Article

Evaluation of the Effect of Comprehensive and Targeted Surveillance on Nosocomial Infections in Nephrology Patients

Jiali Zheng,[1] Jiuying Fei,[2] Hongbo Li,[3] and Yan Xu [1]

[1]*Department of Nephrology, People's Hospital of Dongxihu District, Wuhan, Hubei 430040, China*
[2]*General Practitioner, Jiangjun Road Health Center, Wuhan, Hubei 430040, China*
[3]*Department of Radiology, Wuhan Sixth Hospital, Affiliated Hospital of Jianghan University, Wuhan, Hubei 430015, China*

Correspondence should be addressed to Yan Xu; 2011010224@st.btbu.edu.cn

The article summarizes the control strategy by discussing the risk factors of nosocomial infections in the nephrology department. A survey of hospitalized patients from January 2020 to December 2020 showed that there are six types of bacteria that can cause infections. The age of the patient, the risk of invasive surgery, the low use of antibiotics, and age are all independent factors that affect the risk of nosocomial infections in the patient. The more antibiotics used, the better the infection prevention effect. Among the many risk factors for patient infection, bacterial infection is the main risk factor. *Klebsiella pneumoniae* infection rate was the highest, 33.98%; *Staphylococcus aureus* infection rate was the lowest, 6.80%. Therefore, the nephrology department should focus on strengthening the prevention of *Klebsiella pneumoniae* infection, and implement early prevention and management interventions for various risk factors.

1. Introduction

There are many factors for nephrology patients to have an infection during hospitalization. According to the infection situation during hospitalization, combined with the patient's actual body resistance and immunity level, reasonable analysis of the possible problems of the patient's pathogen effectively control and adjust according to external and internal standards, and control and adjust patients' infection problems. It is possible to strengthen the protection and control process, control the influencing factors of exogenous infection, strengthen the internal management of the hospital environment, and improve the overall quality control layout of the hospital. With the continuous improvement of the hospital's development level, the management requirements for the hospital's information construction also continue to increase. Nowadays, hospital reliability rating has become an important standard to measure the management level of hospitals. The development and reform of hospitals are inseparable from the

informatization construction of hospitals. In the years of development, the informatization construction of Chinese hospitals has made great progress and has a certain scale. But compared with foreign countries, , there is a big gap and also many problems. Nowadays, the main task of constructing informatization is to solve problems and narrow the gap, which still needs improvement in some aspects. This will be conducive for the construction of hospital information in our country, and it is also conducive for the modernization of hospital management in our country. With the advancement of modern medicine, effective management of nosocomial infections has become an important task of hospitals. According to statistics, the probability of nosocomial infections in hospitalized patients reaches 10% each year, and the economic loss is estimated to reach 20 billion yuan. In particular, nosocomial infections will also threaten the lives of other patients in the hospital. Nephrology patients have low self-immunity, the probability of infection increases significantly during hospitalization, and the patients are older and most of

the patients have underlying kidney diseases and other risk factors related to infection, which lead to nephrology inpatients becoming a high-risk group of nosocomial infections.

As part of hospital informatization construction, hospital monitoring is also part of medical quality management [1]. Hospital infection monitoring requires statistical analysis of a large amount of medical data, and the results are complex. In the past, manual exploration and statistical processing were used, but this search method was incomplete and extremely inefficient. It is especially important that real-time monitoring is impossible, especially for the analysis and evaluation of the rational use of antibacterial drugs in hospitals. Because there are many factors involved, it is very complicated and can only be done manually. The analysis did not give the expected results of the hospital, so hospital infection control workers are looking for a set of scientific and practical real-time monitoring software for hospital infection to solve these problems [2]. Hospital infection control includes analysis of the rational use of antibacterial drugs, hospital disinfection and hygiene monitoring, epidemic monitoring, etc. This is a professional medical problem involving multiple disciplines and experts. For the above reasons, the independent development of hospital monitoring software may be a huge and complex task. Therefore, we will continue to study the mainstream hospital sensory monitoring software in the market. The software needs to be compatible with our existing HIS, LIS, seamless connection and data integration, easy-to-use interface, and good human-computer interaction interface.

2. Related Work

In the comprehensive and targeted surveillance of hospital infection of patients, domestic and foreign experts have also conducted many studies. Osman MA's research addresses the existing global nephrology workforce and training capabilities. 125 UN member states responded to the entire survey, of which 121 countries responded to survey questions related to the nephrology labor force. The global density of nephrologists is 8.83 per million population (PMP), and the density of nephrologists reported in high-income countries is 28.52 PMP [3]. Saint believes that preventing medical infections remains an international priority. Although important progress has been made in preventing nosocomial infections in the past few decades, thorny issues still exist, including how to continue to improve hand hygiene and further reduce catheter-related urinary tract infections and other related complications. Perhaps, we should also be looking for "cross-innovation," which means breakthrough discoveries can be made through research in often divergent disciplines [4]. Kim believes that confined spaces appear in medical staff spaces, outpatient spaces, and ward spaces. Since more concentrated spaces in the hospital layout will lead to higher infection rates, it is recommended to plan these spaces separately. The open space of the hospital will be planned as an independent space [5]. In order to identify bacteria in wound infections, Hao S designed an electronic nose system with a sensor array composed of 34 sensors. Eight samples of cultured *Escherichia coli*, *Staphylococcus aureus*,

Pseudomonas aeruginosa, and their mixtures of different concentrations were detected. Using a vector machine as a classifier, no sensor array optimization is required, and the recognition rate is as high as 86.54% [6]. Most patients in the Wagener M study were proven to have secondary bacterial infections, but debridement and subsequent wound management were considered the main treatment. Common microorganisms are Enterobacteriaceae and *Enterococcus*. Antibiotics seem to play a role in the treatment of these patients, so a good antibiotic policy is strongly advocated [7]. AP C's research results indicate that liposome-encapsulated phage cocktails have great potential in the treatment of Klebsiella pneumoniae–mediated infections. Therefore, this strategy can be used as an effective method to treat wound infections caused by *Klebsiella* in individuals who do not respond to conventional antibiotic therapy [8]. These studies have provided a lot of evidence for our experiments, but due to the short research time, there are some doubts about the tested samples, so the test results need to be confirmed by everyone.

3. System Architecture Outline Design

3.1. System Overview. As an important part of the construction of hospital informatization, the hospital monitoring and management information system, it includes several functional modules of subsystems such as system settings, data entry, data query, hospital statistics, suspected case detection, and infection early warning. The main functions of each subsystem are described as follows [9–11].

3.1.1. System Setting Subsystem. The hospital sensor system installation includes the installation and management of various basic hospital sensor software dictionaries. The standard dictionary is expected to improve the work efficiency of operators and reduce various errors caused by manual reference. The dictionary needs to be set in detail and comprehensively. When installing and managing user rights, it can set and change user passwords, backup, retrieve, and initialize data according to user categories.

3.1.2. Data Entry Subsystem. The data entry module can enter infection case data, determine the number of discharged patients, and manually enter the incidence rate. The data entry module can enter infection case data, determine the number of discharged patients, manually enter morbidity, surgical infection cases related to hospital work management, ICU patient infection monitoring data, neonatal risk monitoring data, antimicrobial drug monitoring data, needle stick injury monitoring data, environmental cleanliness monitoring data, and case report card records. This function is to facilitate the manual input of data into the computer for statistics.

3.1.3. Data Query Subsystem. According to different combinations of situations, the data query module queries recorded data, infection incidence, surgical infection, patient infection, high-risk newborns, antibacterial drugs, needle

stick injuries, and environmental health data from different angles.

3.1.4. The Statistics Subsystem of the Hospital.
Statistics of infection incidence, surgical infection, patient infection, neonatal risk, antibacterial drugs, acupuncture wounds, and environmental health data were calculated according to the departments.

3.1.5. Suspected Case Search, Infection Outbreak Early Warning Subsystem.
It can monitor the overall dynamics of the hospital in real time, with detection and early warning functions, to achieve early warning and reminders of the same infection case in the same department within a period of time [12].

3.2. Work Investigation of Hospital Infection Management Information System.
After many reviews, comparisons, and analyses, combined with the actual needs of the hospital infection control workers, the hospital infection control information system used by the hospital infection control workers and the information center staff is determined. Through communication with the staff of the hospital infection department, there are 3000 cases of nosocomial infections, and the number of nosocomial infection cases per year is counted, as shown in Figure 1.

It can be seen from Figure 1 that as the number of hospital beds increases, hospital infections also increase. Traditional manual search, statistics, and analysis methods are inefficient and have high error rates. Starting from the prevention and treatment of nosocomial infections, a nosocomial infection management information system was launched. This can not only solve the above problems but also pull data from different angles, draw related work plans, and meet the business needs of various related departments. The business report of the infection control system is shown in Figure 2. The use case diagram of the system function is shown in Figure 3 [13].

As the staff of the hospital information center, you need to understand the hospital infection prevention and treatment business process. This facilitates communication with business departments, satisfies their needs, and aligns the software with actual business processes. After communicating with hospital staff by visiting similar software usage conditions in other hospitals, according to the specific situation, improve the specific business process of the hospital's work. Especially, report the business process of the infection control system. If clinicians find that a patient has a nosocomial infection, they must promptly notify the nosocomial infection control department, and report the nosocomial infection to the higher-level department after the diagnosis.

The design of hospital infection science information system and the analysis of user needs are mainly realized by UML tools. It is a language used to visually model software-intensive systems. As a model language, UML allows software developers to focus on creating product models and structures. Most models are represented by diagrams. A typical modeling diagram usually consists of several blocks or boxes, contact lines and text, which are used as additional information for the model. The communication with users in the initial stage of the system is to determine the main operating modules according to the main business process.

The functional architecture of the system is represented by the basic ideas of "big to small" and "up to down" attenuation similar to structural analysis. The functional architecture of the system is an important software infrastructure. Software developers can easily visualize the functions of each module in the system through a clear system function architecture. In the development and design process, you can also perceive the function of each module according to the functional architecture diagram, and understand the relationship between the functional modules. The system function frame diagram is shown in Figure 4 [14].

The state diagram of the system describes the dynamic behavior of an entity based on event reaction, and shows how the entity reacts to different events according to the current state. This enables the events and conditions of the object to reach these states, as well as the actions that occur when these states are reached. As shown in Figure 5, it is the state diagram reported by the hospital sensor card of the hospital sensor management system [15].

3.3. System Database Design.
Basic department information table: This table is used to store the basic information of hospital departments. Most of the units and reports of the system are based on the department name. There is no correct department setting, so many units will make mistakes or produce no results. When the basic department information table is called, it is called with officeI (department number) and office (department name) as the primary keys. Table 1 is the basic department information table [16].

Patient medical order information table: The information in the patient medical order information table is called from the HIS system through the interface. This report contains almost all the patient's medical order information. Identify and locate the patient by hospitalization ID and hospitalization number, and identify and locate the doctor's order by the doctor's order name and doctor's order number. The patient's medical order information form is mainly used for the hospital clinic to judge whether the patient is really a hospital-infected patient based on the medical advice information when reviewing the hospital's report card, and as the basis for the approval of the review. The name of the executive department and the number of the executive department are based on the statistical reports of each executive department. If it cannot be extracted from HIS, it will affect the formation of many reports. And, if the prescribed doctor cannot be extracted, the report of the doctor's statistics cannot be generated. The patient's medical order information form provides comprehensive information
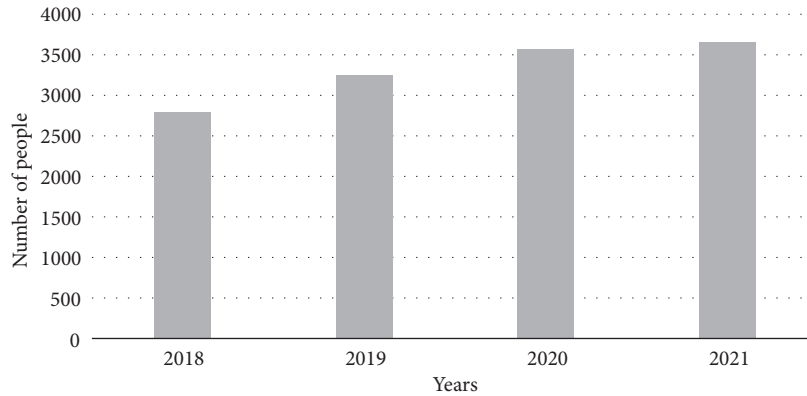
FIGURE 1: The number of infections in the hospital each year.
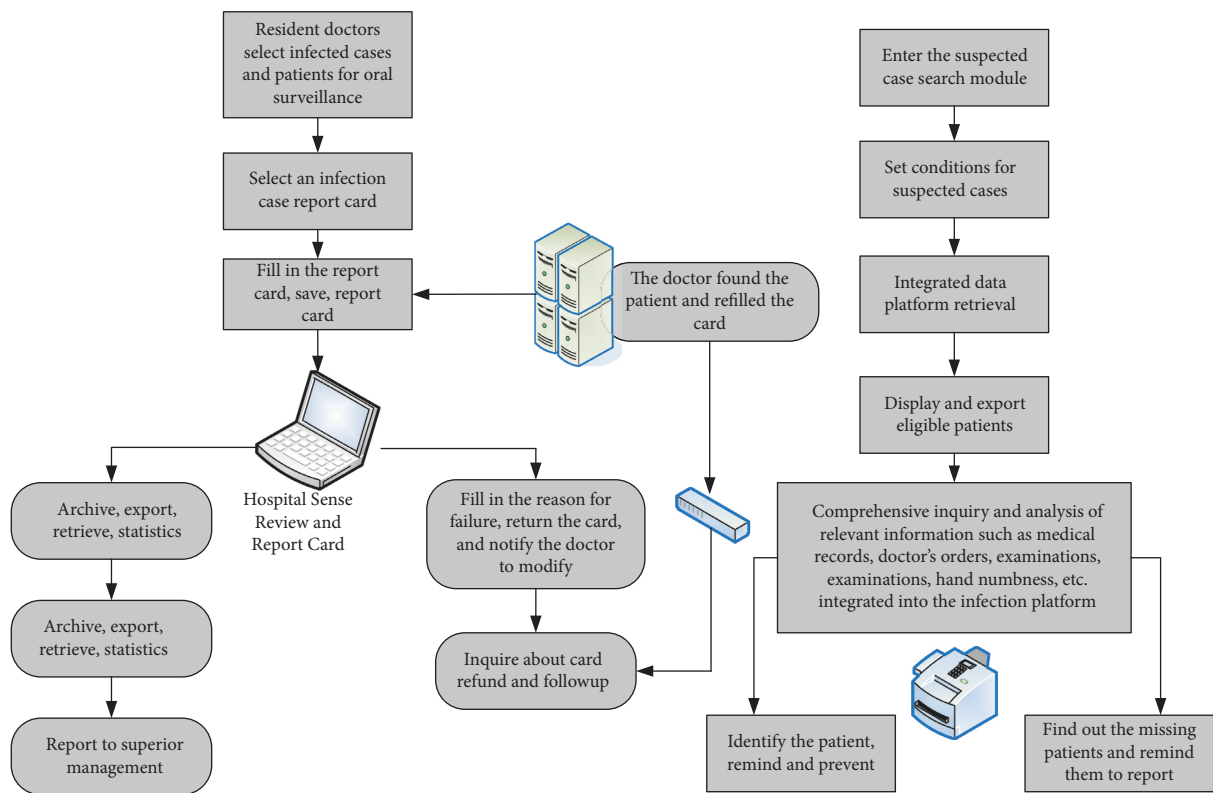
FIGURE 2: Infection management system reporting business process and suspected case search management business process.

for auxiliary diagnosis. It has a high degree of integration, displays the original medical order in the medical order system, and provides data support for the follow-up statistical reports of the hospital sensory department from various angles. It can also more comprehensively support the inquiries, analysis, and statistics of the hospital sensor system. Table 2 is the patient's medical order information table.

3.4. K-Means Algorithm. The algorithm takes k as an input parameter and divides the object set into k clusters. The average value of objects in a cluster is measured by the

uniformity of the cluster. These average values are considered to be the center of gravity of the cluster, or become the center of mass. We can use the squared error criterion to determine the criterion function, which is defined as follows:

$$E = \sum_{i=1}^{k} \sum_{p \in Ci} |p - m_i|^2. \tag{1}$$

According to the different dimensions of the distance between classes, different systematic clustering methods can be carried out. There are many methods of systematic
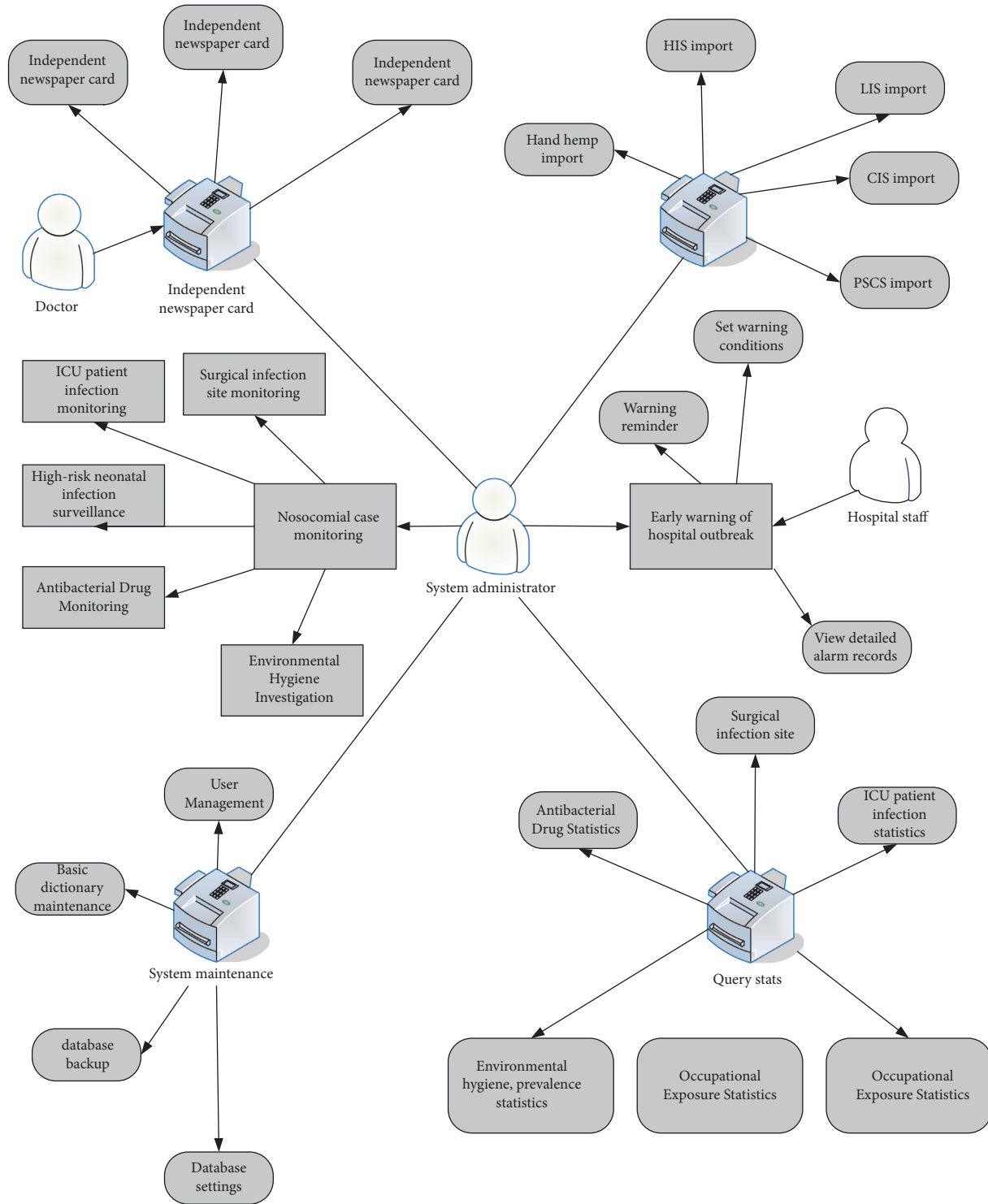
Figure 3: System Sichuan example map.

clustering, among which the most commonly used are: longest distance method, class average method, shortest distance method, intermediate distance method, variable class average method, variable method, deviation square and center gravity method, and so on. The steps of the cluster analysis system method are the same.
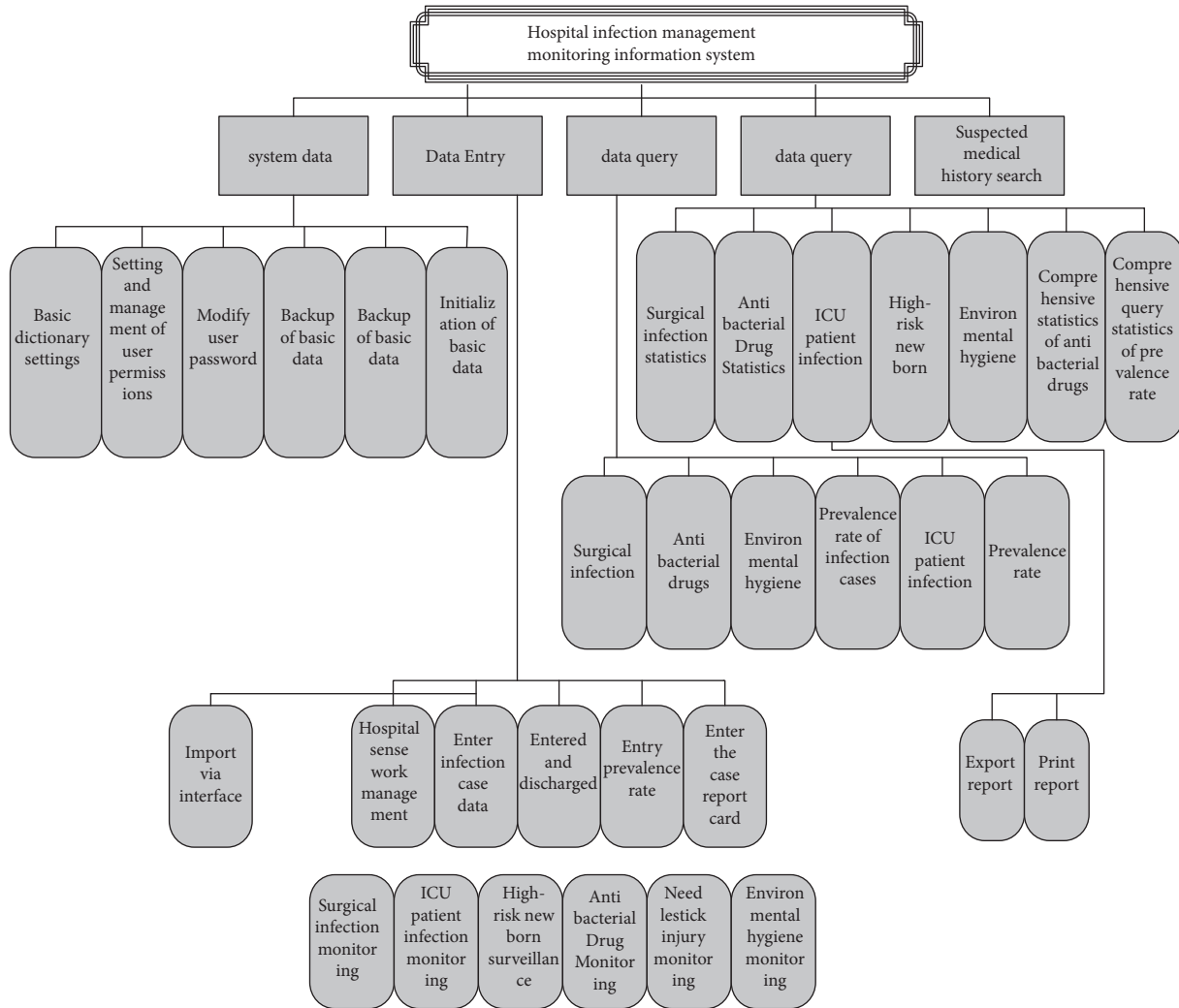
FIGURE 4: System functional architecture diagram.

Define the distance between the class Gi and Gj as the distance between the two nearest samples, namely:

$$D_{ij} = \min_{Gi \in Gi \cdot Gj \in Gj} d_{ij}. \tag{2}$$

Let the classes Gp and Gq be merged into a new category, denoted as Gr. Then, the distance between any type of Gk and Gr is:

$$Dkr = \min_{Gi \in Gi \cdot Gj \in Gj} d_{ij},$$

$$\min \left\{ \min_{xi \in Gk \cdot x j \in Gp} d_{ij}, \min_{xi \in Gk \cdot x j \in Gq} d_{ij} \right\}, \tag{3}$$

$$\min \{ D_{kp}, D_{kq} \}.$$

The dimension and magnitude of each attribute will be different. Therefore, some very large properties can be

found in this specific calculation process, which will affect the distribution. Therefore, these quality data need to be standardized, and each attribute value is combined into a specific standard, so that the combined data have a specific set of common numerical characteristics. The method of unifying the data can be done like shown below:

$$x_{ij} = \frac{x_{ij} - \overline{x}_j}{S_j}, \tag{4}$$

$$x_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}, \quad S_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(x_{ij} - \overline{x}_j \right)^2}. \tag{5}$$

Here, the following extreme value standardization formula can be used to calculate:
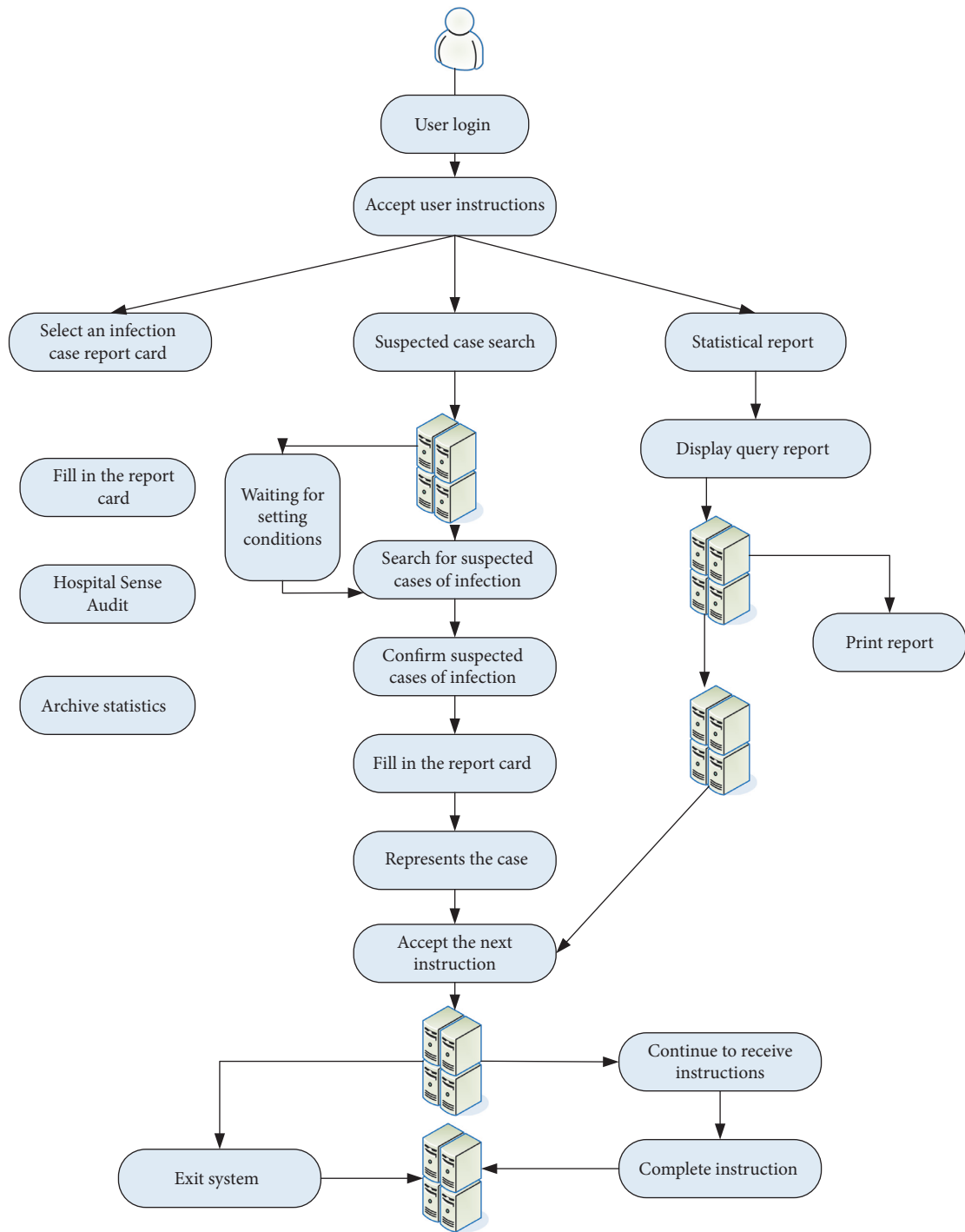
FIGURE 5: The hospital sensor card system status diagram is reported.

TABLE 1: Basic information of the departments of our hospital.

Serial number	Column name	Type of data	Length	Allow empty	Instruction
1	OfficeID	char	10	NOT NULL	Department number
2	Office	verchar	30		Department name
3	Officekindid	char	10		Standard department number
4	Ifdelete	bit	1	NOT NULL	Delete or not
5	Memoryid	char	10		Mnemonic
6	Standofficeid	char	10		Standard department
7	Ifcaseoffice	bit	1		Whether the case monitoring department
8	Ificrcumoffice	bit	1		Whether the environmental monitoring department
9	Ificu	bit	1		Whether ICU
10	Ifbak	bit	1	NOT NULL	Is it commonly used

TABLE 2: Patient's medical order information form.

Serial number	Column name	Type of data	Length	Allow empty	Instruction
1	ZYID	Vaechar	50	NOT NULL	Unique ID number for hospitalization
2	Patient id	Vaechar	50	NOT NULL	Hospital number
3	Visit id	Int	8		Number of hospitalizations
4	name	Vaechar	50	NOT NULL	Name
5	Flag	Int	4		Long testimony
6	Start date	Datetime	8		Open date
7	Stop date	Datetime	8		Date of suspension
8	Order days	Numeric	8		Days of medication
9	Order class	Vaechar	50		Medical order item classification name
10	Order code	Vaechar	16	NOT NULL	Order number
11	Order name	Vaechar	200	NOT NULL	Name of doctor's order
12	Order speci	Vaechar	50		Drug specifications
13	YCJL	Numeric	8		Dose
14	Dosage units	Vaechar	20		unit
15	YCSL	Numeric	8		Quantity
16	SYPC	Vaechar	50		Frequency of use
17	Usemaediway	Vaechar	50		Purpose of administration
18	Uspurpose	Vaechar	50		unit price
19	YPDJ	Numeric	8		Remark
20	Order memo	Vaechar	100		Times per day
21	MRCS	Int	1		Daily dosage
22	Qty day	Numeric	8		Total amount
23	Qty sum	Numeric	8		Executive department number
24	Execoffice code	Vaechar	20	NOT NULL	Executive department name
25	Execoffice name	Vaechar	50	NOT NULL	Prescribe doctor number
26	Bdoctor code	Vaechar	20		Name of prescribing doctor
27	Bdoctor name	Vaechar	20		Stop ordering doctor number
28	Edoctor code	Vaechar	20		Name of doctor who stopped advising
29	Edoctor name	Vaechar	20		Antibacterial drugs
30	Flag kjyw	bit	1		Antibacterial drugs

$$x_{ij}'' = \frac{x_{ij}' - x_{j\,min}'}{x_{j\,max}' - x_{j\,max}'}, \tag{6}$$

$$R = \begin{bmatrix} r11 & r12 & \cdots & r1n \\ r21 & r22 & \cdots & r2n \\ \cdots & \cdots & \cdots & \cdots \\ rn1 & rn2 & \cdots & rnm \end{bmatrix}. \tag{7}$$

There are many commonly used methods to calculate rij, such as the distance method, geometric mean minimum method, correlation coefficient method, and angle cosine method. The Euclidean distance method is used here, and the distance coefficient is defined as d_{ij}, and the D matrix is:

$$D = \begin{bmatrix} d11 & d12 & \cdots & d1n \\ d21 & d22 & \cdots & d2n \\ \cdots & \cdots & \cdots & \cdots \\ dn1 & dn2 & \cdots & dnm \end{bmatrix}, \tag{8}$$

$$d_{ij} = \sqrt{\sum_{k=1}^{n} \left(x_{ik} - x_{jk} \right)^2}. \tag{9}$$

Now take the first indicator as an example for data normalization:

Average is:

$$\overline{x}_i = \frac{1}{n} \sum_{i=1}^{n} x_{i1}$$

$$= \frac{(99.51 + 99.16 + 93.12 + \cdots + 80 + 78.67)}{23} \tag{10}$$

$$= 95.41.$$

The standard deviation is:

$$S1 = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(x_{i1} - \overline{x}_1 \right)^2}$$

$$= 6.47. \tag{11}$$

According to (4), to get the standard value:

$$x_{11}' = \frac{x_{11} - \overline{x}_1}{S1}$$

$$= \frac{(99.51 - 95.41)}{6.47}$$

$$= 0.633,$$

$$x_{21}' = \frac{x_{21} - \overline{x}_1}{S1}$$

$$= \frac{(99.61 - 95.41)}{6.47}$$

$$= 0.579. \tag{12}$$

According to (6), the standardized data are compressed into the closed interval [0, 1]:

$$x_{11}'' = \frac{x_{11}' - x_{min\,1}'}{x_{max\,1}' - x_{min\,1}'}$$

$$= \frac{(0.633 + 2.587)}{(0.709 + 2.587)}$$

$$= 0.977,$$

$$x_{21}'' = \frac{x_{21}' - x_{min\,1}'}{x_{max\,1}' - x_{min\,1}'}$$

$$= \frac{(0.579 + 2.587)}{(0.709 + 2.587)}$$

$$= 0.96. \tag{13}$$

Setting is to establish the fuzzy similarity relationship, where the fuzzy similarity relationship between various departments is established. According to (7) and (8), the similarity relationship is calculated with the first row and second row of the matrix as examples.

$$d_{12} = \sqrt{\sum_{k=1}^{n} \left(x_{1k} - x_{2k} \right)^2}, \tag{14}$$

$$\sqrt{\left(x_{11} - x_{21} \right)^2 + \left(x_{12} - x_{22} \right)^2 + \cdots + \left(x_{1n} - x_{2n} \right)^2} = 2.072.$$

4. Objects and Methods of Comprehensive Research on Infection

4.1. The Object. Surgery departments usually include cardiothoracic surgery, general surgery, hepatobiliary surgery, neurosurgery, trauma and orthopedics, otolaryngology, urology, obstetrics and gynecology, hand and foot surgery, and burns and plastic surgery. Most of the inpatients in the hospital come from the province. The survey selected all surgical patients who were hospitalized for more than 48 hours from January 1, 2020 to December 31, 2020.

4.2. Exclusion Criteria. Excluding surgical patients who have not been hospitalized for more than 48 hours or who have suffered nosocomial infections before the operation shall be implemented in accordance with the regulations promulgated by the Ministry of Health. Nosocomial infections refer to infections that occur in hospitalized patients, including infections that occur during hospitalization and infections that occur in hospitals after discharge. Infections among hospital staff are also related to nosocomial infections but do not include infections before admission and infections that already existed at the time of admission.

4.3. Experimental Method. It collects basic information about patients undergoing surgery in the hospital, including gender, age, medical record number, department, date of admission, and basic information about surgery (days before surgery, emergency/elective schedule, anesthesia, type of surgery, operation time, etc.). The main purpose of antibiotics is whether to use antibiotics preventively, whether to use antibiotics before surgery, where and when to use antibiotics before surgery, etc. [17].

The study subjects were observed until the onset of nosocomial infections, discharge, or death. Surgery patients with nosocomial infections were included in the case group, and patients without infection during the operation were included in the control group. For surgical patients with nosocomial infections, use the nosocomial infection case questionnaire to collect more information about the patient's infection: date of infection, location of infection, etiological examination results, drug susceptibility test results, etc. According to descriptive research methods, calculate the morbidity, morbidity intensity, underreporting rate, standardized disease intensity, and other indicators of surgical patients, and calculate the distribution of hospital infection sites and departments of surgical patients. It also calculates the risk index of surgical patients. After adjusting the risk index, it calculates the morbidity and disease density of surgical patients [18].

$$\text{Incidence} = \frac{\text{Number of new hospital infections during the same period}}{\text{Number of hospitalized patients monitored during a certain period}} \times 100\%,$$

$$\text{Incidence rate} = \frac{\text{Number of new hospital infections during the same period}}{\text{Number of hospitalized patients monitored during a certain period}} \times 100\%,$$

$$\text{False negative rate} = \frac{\text{Number of underreported nosocomial infections during the same period}}{\text{The actual number of infections in a certain period}} \times 100\%,$$

$$\text{Incidence density} = \frac{\text{Number of new hospital infections during the same period}}{\text{Number of bed days of inpatients monitored during a certain period}} \times 100\%.$$

(15)

The quantitative data of the study were analyzed by *t*-test or nonparametric test, case management research method was used, and the difference between the case group and the control group was analyzed by chi-square test or Fisher's exact probability method. Logistic regression analysis was performed on the research factors and the nosocomial infection factors, and the OR value of related research factors was calculated. The variables with $P < 0.1$ in the above single factor analysis were included in the multiple logistic regression model for analysis. The study uses SPSS21.0 software to analyze the relevant data. All hypothesis tests are carried out by two-sided tests, with $\alpha < 0.05$ as the significance level for judgment.

In the design stage, the design of this project has been studied and demonstrated by many experts and professors to ensure that the project design is scientific, reasonable, and feasible. Pre-survey the research objects and continuously improve the content of the questionnaire. In the project implementation stage, the investigators are trained before conducting the investigation to ensure that the investigators can clarify the purpose of the investigation, and are familiar with the content, methods, and techniques of the investigation. During the investigation, the nosocomial infection staff checked the questionnaire in time, found the wrong items and missing items, and made supplements and improvements in time. Proactively communicate with clinicians, nurses, microbiological testers, patients, or accompanying personnel in time to ensure the high efficiency of the investigation. In the data analysis stage, communicate with statistical professionals in time, find problems, and solve them in time [19]. Patients with nephropathy are mostly chronic kidney disease. Due to the

(a)

(b)
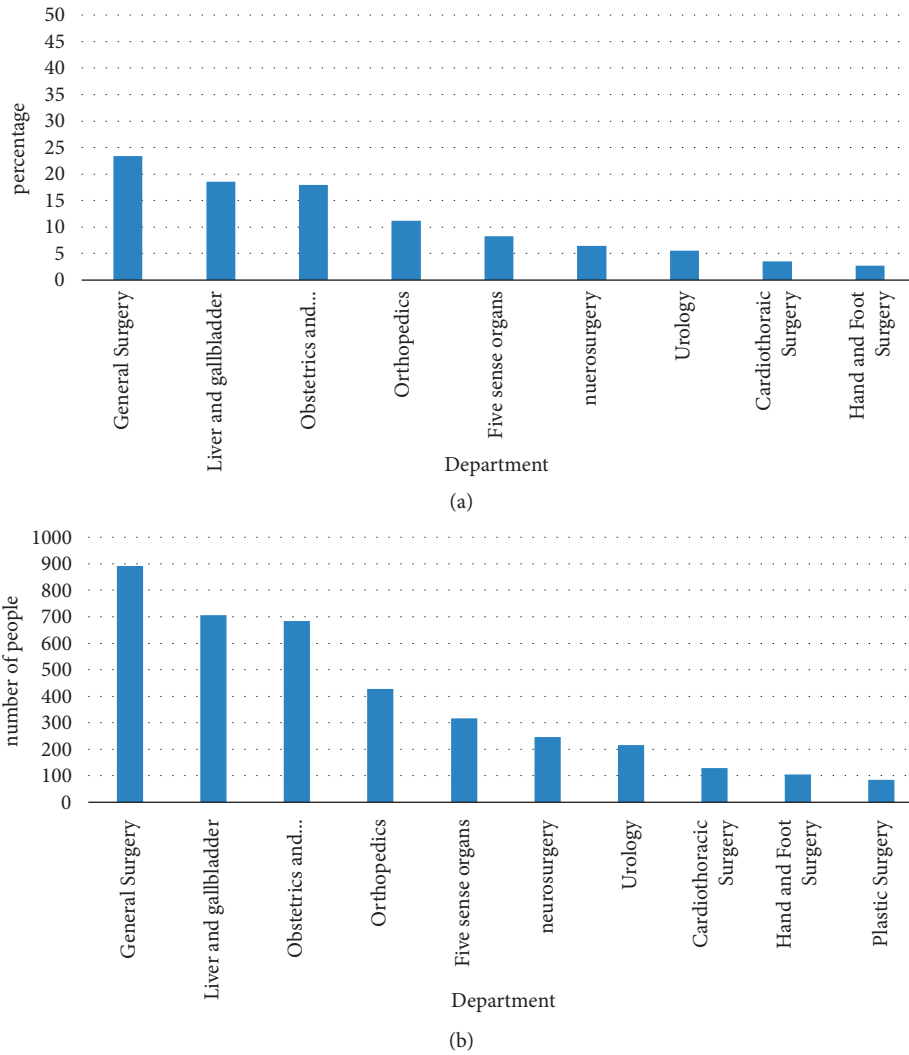
Figure 6: Distribution table of surgical patients in each department.

Table 3: The use of antibacterial drugs and preventive antibacterial drugs in surgical patients.

Department	Number of operations	Number of users	Utilization rate (%)	Preventive medication	Composition ratio (%)
Cardiothoracic surgery	134	116	86.6	94	81.4
Neurosurgery	247	203	82.2	182	89.8
Traumatology	427	341	79.9	282	82.7
General surgery	891	615	69.0	467	75.9
Hepatobiliary surgery	706	510	72.2	304	59.7
Urology	210	175	83.3	149	85.3
Hand and foot surgery	104	89	85.6	73	81.9
Five sense organs	316	164	51.9	81	49.6
Plastic surgery	85	66	77.6	21	31.5
Obstetrics and gynecology	683	427	62.5	225	52.8
Total	3804	2706	71.7	1880	69.5

long course of disease, the increased permeability of the glomerular basement membrane, and the decrease in glomerular filtration, the patient's own albumin loss is relatively large, and the patient's immune ability is low, and it is easy to be susceptible. Infection caused by pathogenic bacteria ultimately affects the clinical treatment of the primary disease, aggravates the patient's condition, and even endangers the patient's life.

4.4. Research Situation. A total of 3804 surgical patients were investigated this time: 1741 were males, accounting for 45.7%; 2063 were females, accounting for 54.3%. There were 221 surgical patients aged 14 and under, accounting for 5.8%; 3085 surgical patients aged 15–59, accounting for 81.1%; 498 surgical patients aged 60 and above, accounting for 13.1%. The oldest is 89 years old and the youngest is 2 years old. The average age is 38.1 ± 13.9 years, of which the average age of men is 39.6 ± 13.8 years, and the average age of women is 35.1 ± 14.7 years [20]. Among all the patients investigated, 891 were in general surgery, 706 were in hepatobiliary surgery, 683 were in obstetrics and gynecology, 426 were in trauma and orthopedics, 316 were in ENT, and 247 were in neurosurgery, 210 cases of urology, 135 cases of cardiothoracic surgery, 104 cases of hand and foot surgery, and 85 cases of burn plastic surgery department. Figure 6 shows the distribution of surgical patients in each department.

A total of 3804 surgical patients were investigated. The investigation and analysis of the use of antibacterial drugs showed that 2706 surgical patients used antibacterial drugs, and the use rate of antibacterial drugs was 71.1%. The use of antibiotics in cardiothoracic surgery, neurosurgery, urology, and foot surgery is relatively high (see Table 3 for antibiotic use in patients in various departments). Among all surgical patients who used antibacterial drugs, there were 2015 surgical patients using antibacterial drugs in combination, accounting for 74.5%; 593 surgical patients using dual antibacterial drugs, accounting for 74.5%; 593 surgical patients using dual antibacterial drugs, accounting for 21.9%; and 98 surgical patients using triple antibacterial drugs, accounting for 3.6% [21].

During the operation, there were 1880 surgical patients who used antimicrobial prophylactically, accounting for 49.4% of all surgical patients; 1037 cases of preoperative preventive medication, accounting for 55.2% of patients using prophylactic antibiotics during surgery; and 1496 cases of postoperative preventive medication, accounting for 79.6% of patients with surgical preventive medication. There were 697 patients who used antibiotics before and after surgery, accounting for 37.1% of the patients who used antibiotics preventively. The use of preventive antibiotics in neurosurgery and urology surgery is relatively high (see Table 3 for the use of preventive antibiotics in patients undergoing surgery in different departments). The average time to start medication before surgery was 36.3 hours before surgery, and the average time to medication after surgery was 4 days.

This survey shows that the most commonly used antibacterial drugs are cephalosporins, which account for 40.31% of all research subjects using antibacterial drugs; followed by penicillin antibacterial drugs, accounting for 16.95% of all antibacterial drugs; aminoglycosides, accounting for 12.54%; quinolones, accounting for 11.18%; and metronidazole/tinidazolem accounting for 8.4%. Figure 7 shows the distribution of antibacterial drugs in surgical patients [22].

Among all the surgical patients, there were 3265 elective surgery patients and 539 emergency surgery patients. There were 1685 patients undergoing general anesthesia and 2119 patients undergoing local anesthesia. There were 1694 patients with type I incision surgery, 1958 patients with type II incision surgery, and 152 patients with type III incision surgery. The average hospital stay before operation was 6.3 days, the average operation time was 118 minutes, and the average hospital stay was 14.2 days. In this survey, there were 1,050 patients who did not undergo any invasive operation, accounting for 27.6%; the number of patients who had undergone invasive operation was 2,754, accounting for 72.6%. Among them, the use rate of drainage tube accounted for 49.6%, the use rate of urinary catheter accounted for 28.7%, arteriovenous intubation accounted for 11.4%, and the remaining invasive procedures accounted for 10.3%. There were 1077 patients who had undergone more than two invasive procedures, accounting for 39.1%.

Among all surgical patients investigated, the total number of observation bed days was 32714 bed days, the number of hospital infection cases was 219, and the incidence rate of surgical patients was 5.8%; the number of infection cases was 236, and the incidence rate was 6.2%. A total of 206 cases of nosocomial infection were reported by surgical clinicians, and the underreporting rate was 12.7%.

The incidence of nosocomial infections fluctuated between 4.5% and 6.5% between January and December 2020. As shown in Figure 8, there was no statistically significant difference in the incidence of each month. The incidence density of surgical patients was 6.7‰, and the monthly incidence density fluctuated between 5.9‰ and 7.3‰. As shown in Figure 8, there was no statistically significant difference in incidence density between months. In the case group, there are 143 males, accounting for 65%; 76 females, accounting for 35%. The severity of nosocomial infections is shown in Table 4. The gender differences in nosocomial infections of surgical patients are statistically significant. Among them, the average age of men is 40.5 ± 14.1, and the average age of women is 34.9 ± 13.2 years. The incidence of nosocomial infection in each age group is shown in Table 4. The age difference of nosocomial infections among surgical patients was statistically significant [23].

In the case group, nosocomial infections of patients undergoing operations in different clinical departments are mostly primary respiratory infections, but surgical infections are more common in the Department of Traumatology and Orthopedics and Hand and Foot Surgery. The structure of nosocomial infections in each case is arranged as follows: lower respiratory tract infection, surgical site infection, upper respiratory tract infection, urinary tract infection, and gastrointestinal tract, skin, and soft tissue infections. As shown in Figure 5. The lower respiratory tract is the most common site of infection, followed by the surgical site and the upper respiratory tract. Nosocomial infections of surgical patients ranked in the top 3 infections, the sum of which exceeded 80%. The incidence of lower respiratory tract infection was 2.1%, and the incidence density was 2.5‰; the incidence of the surgical site was 1.7%, and the incidence density was 2.0‰; the incidence of upper respiratory tract infection was 0.9%, and the incidence density was 1.0‰. The incidence of nosocomial infections in other parts of the body decreased successively. Table 5.
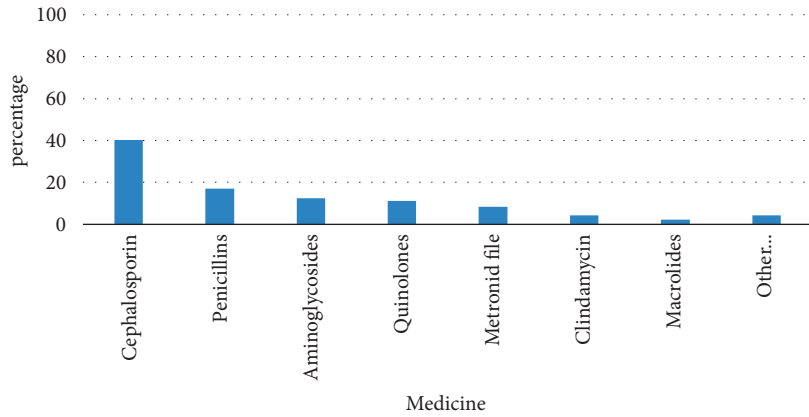
FIGURE 7: Distribution of antibacterial drugs in surgical patients.

(a)

(b)
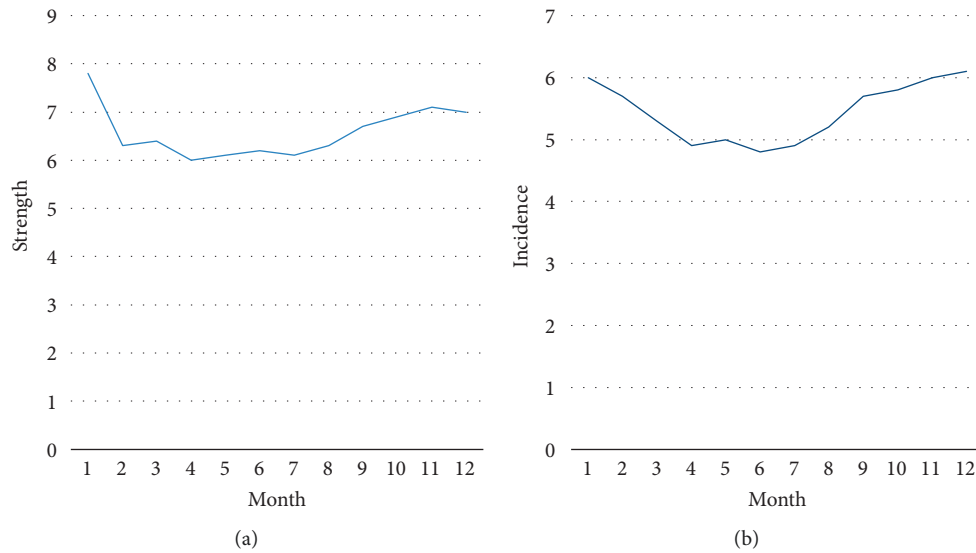
FIGURE 8: Intensity of nosocomial infections in each month.

TABLE 4: The incidence of nosocomial infections by gender and age in the case group.

	Grouping	Number of hospital infections	Incidence	Incidence density	X^2	P Value
Gender	Male	143	8.2	9.5		
	Female	76	3.7	4.3	7.81	0.005
Age	15–59 years old	145	4.7	5.1		
	≤14 years old	15	6.8	5.9	47.56	0.001
	≥60 years old	59	11.8	11.4		

First, conduct a gender and age balance test on the grouping of relevant research factors in the case group and the control group to eliminate the interference of the two confounding factors of gender and age. The results showed that there was no statistical difference between the groups; X^2 test and factor Logistic regression were used to analyze the relevant factors of the case group and the control group again to investigate the risk factors. In the analysis, the first group in the grouping was used as the control group, and the

other groups were compared and analyzed with the first group as the control [24].

The results showed that gender, age, preventive use of antibacterial drugs during surgery, preoperative hospital stay, surgical preparation, anesthesia method, surgical incision type, surgical duration, whether there was invasive surgery, and whether there was a disease were statistically significant ($P < 0.05$). It can be seen from Figure 9 that the incidence of nosocomial infections in male surgical patients

TABLE 5: Incidence rate, incidence density, and composition ratio of nosocomial infections in various parts.

Site of infection	Number of cases	Incidence rate (%)	Incidence density (%)	Composition ratio (%)
Lower respiratory tract	81	2.1	2.5	37.0
Surgical site	65	1.7	2.0	29.7
Upper respiratory tract	34	0.9	1.0	15.5
Urinary tract	20	0.2	0.6	9.1
Gastrointestinal tract	7	0.5	0.2	3.2
Skin and soft tissue	3	0.2	0.1	1.4
Other parts	9	0.1	0.3	4.1
Total	219	6.2	6.9	100
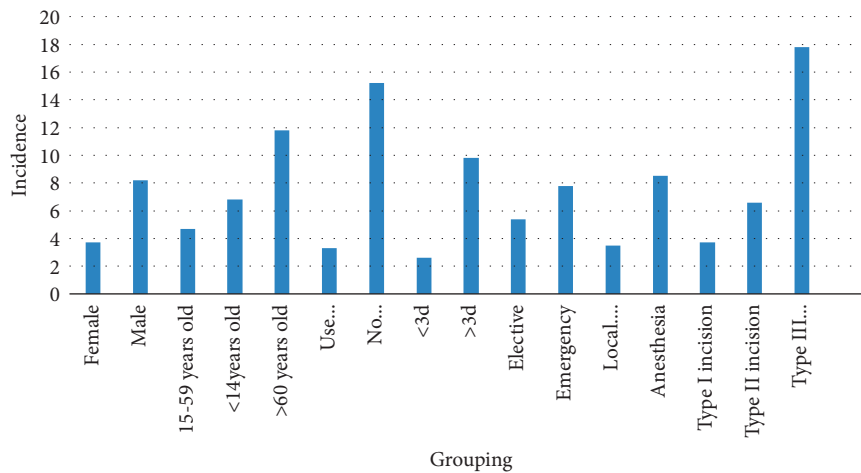
(a)
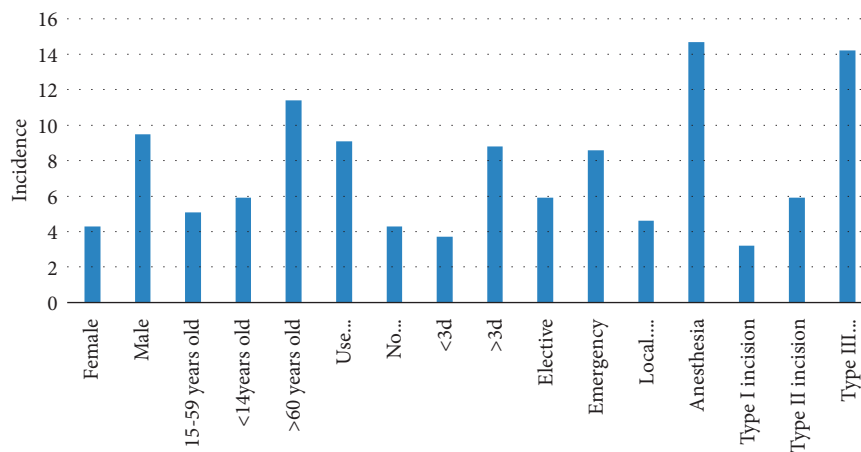
(b)

FIGURE 9: Single factor analysis of factors related to nosocomial infection.

is higher than that of female surgical patients; the incidence of nosocomial infections in surgical patients over 60 years old is higher than that of surgical patients in other age groups, and the incidence of nosocomial infections in surgical patients aged 15–59 is the lowest; the incidence of nosocomial infections in surgical patients who did not use antibacterial drugs during the operation was higher than that of patients who used preventive antibacterial drugs; hospital infections in surgical patients who were hospitalized for more than 3 days significantly exceeded those who were hospitalized for less than 3 days before surgery; the intensity of nosocomial infections in emergency surgery patients is higher than that in elective surgery patients; the nosocomial infection rate of general anesthesia patients is higher than that of local anesthesia patients; the infection intensity of nosocomial infection surgery patients in Class III surgery is significantly higher than other types of surgery patients; and the prevalence of nosocomial infections in type I patients is
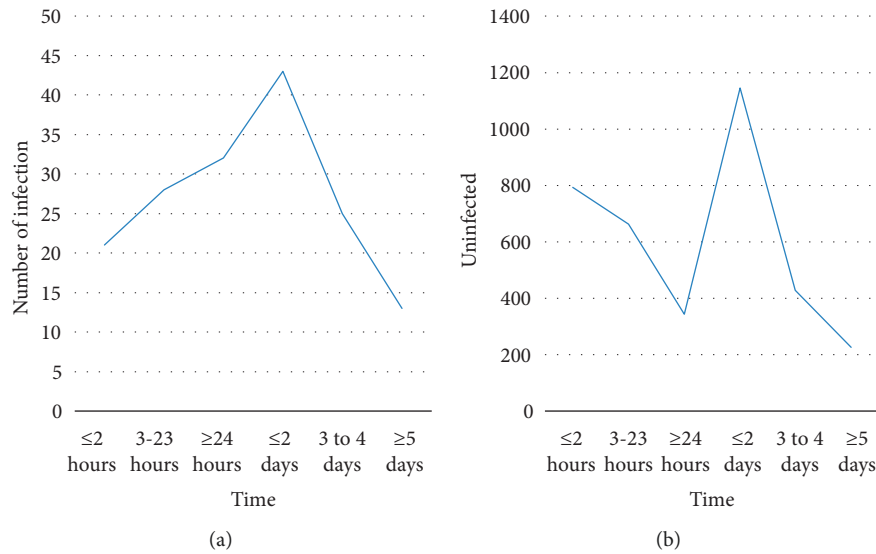
FIGURE 10: The influence of the use of antibacterial drugs during the operation on the intensity of nosocomial infections in surgical patients.

the lowest in surgical fracture operations. The intensity of nosocomial infections in surgical patients increases with the increase of operation time, and the operation time exceeds 6 hours, and the intensity of nosocomial infections in surgical patients is the highest; the nosocomial infection rate of patients with invasive surgery is significantly higher than that of patients with noninvasive surgery; the incidence of nosocomial infections in patients with underlying diseases is higher than that of patients without underlying diseases [25].

From the multivariate logistic regression analysis, it can be seen that the OR 95% confidence interval of whether antibacterial drugs are used during the operation period is 0.957, 1.936, including 1, which may be a risk factor or a protective factor. Therefore, the classification study of whether to use antibacterial drugs during the operation is carried out, and the X^2 test and logistic single factor analysis are carried out on whether to prevent the use of antibacterial drugs, the timing of medication, the course of treatment, and whether to use antibacterial drugs in combination. In the analysis, the first group was used as the control group, and the other groups were compared and analyzed with the first group as the control. The results show that if antibacterial drugs are used prophylactically during surgery, the preoperative antibacterial drug use time and the combined use time of antibacterial drugs are statistically significant; the preventive use of antibacterial drugs before surgery is not statistically significant, as shown in Figure 10. It shows that the preventive use of antibacterial drugs during surgery can reduce the intensity of nosocomial infections during surgery [26]. Preventive use of antibacterial drugs for less than 2 hours before surgery can effectively prevent nosocomial infections in surgical patients. The more antibacterial drugs used, the stronger the intensity of nosocomial infections in surgical patients.

From the multivariate logistic regression, 9 risk factors of nosocomial infection in surgical patients can be obtained. For these 9 risk factors, the risk factor index of each patient is

calculated. Calculation method: The basic risk factor index of surgical patients is 0. When a risk factor appears in a surgical patient, the risk factor index of the surgical patient increases by 1 point until the risk factor index of the surgical patient is calculated. Therefore, each patient's score can be divided into 0–9 points, and the corresponding risk factor index is divided into 10 levels. It can be seen from the results that surgical patients with one risk factor are the most, followed by surgical patients with two risk factors. With the increase of risk factors, the risk factor index gradually increases, the number of hospital infections also increases, and the incidence of surgical patients increases accordingly. Finally, calculate the risk index of each surgical patient to determine the average risk index, and the result is 1.7. The results of the study showed that nosocomial infection in nephrology patients was related to patient age, course of disease, length of hospital stay, diabetes mellitus, invasive procedures, high 24-hour urinary protein, and serum albumin. The older the patient is, the lower the body's autoimmunity is, and the easier it is to be invaded by pathogenic bacteria and complicated by infection; the longer the patient's disease course, the lower the renal function.

5. Discussion

The development of nosocomial infection work is extremely uneven, and the methods used in the nosocomial infection work carried out by major hospitals are uneven. In hospital infection surveillance research, epidemiology provides a series of research methods for it. The commonly used research methods are: cross-sectional studies, case-control studies, retrospective investigations, and prospective investigations, all of which belong to observation methods. Cross-sectional research, also known as current prevalence rate, can calculate hospital infection rate. It is the most commonly used method in descriptive epidemiological research. Most hospital infection surveillance uses this

indicator. Prospective nested case-control research combines the advantages of prospective investigation. The advantage of a prospective nested case-control study is that it starts to collect patient-related data at the beginning of the study, and the selection bias and information bias are small; because the sample size required by the research is smaller than that of prospective investigations, it can save a lot of manpower, material resources, and financial resources; it can calculate the intensity of nosocomial infections; it can detect the trend of nosocomial infection epidemic or outbreak in time. It is a case-control research, and cleverly avoids its shortcomings, meets the requirements of causal research, and demonstrates the characteristics of high causal intensity. Therefore, applying the prospective nested case-control study to the nosocomial infection target surveillance study, the method is appropriate, and it will gradually receive the attention of the majority of medical staff and be widely used. With the continuous introduction of new technologies such as precision instruments and minimally invasive techniques, the damage caused by surgery is getting smaller and smaller, and the incidence of nosocomial infections caused by surgical site infections has decreased significantly. However, the use of new technologies also brings negative effects, such as increased opportunities for invasive operations and continuous updating of clinical drugs, which make bacteria continue to mutate and increase the probability of infection in nonsurgical sites. Therefore, the form of nosocomial infections is not optimistic, and targeted surveillance should be carried out in the daily work of nosocomial infection. Flexible application of epidemiological methods to nosocomial infection work, a prospective nested case-control study of inpatients. It further improves the nosocomial infection monitoring system, ensures the quality of monitoring, reduces the risk factors of nosocomial infection to a minimum, and then reduces the intensity of nosocomial infection [27].

Urinary tract infection is one of the most common complications in clinical patients, and its incidence is second only to respiratory tract infection among various types of nosocomial infections. The high incidence of urinary tract infection may be caused by inappropriate indwelling catheters and frequent use of various interventional diagnostic and therapeutic measures. During the operation phase, the most prone to nosocomial infection is during the operation. The most ideal preventive antibacterial drug should be used during the entire operation. The concentration of antibacterial drugs reaches the proper concentration in the body, kills pathogenic microorganisms in the body, and achieves the purpose of preventing infection. Therefore, the use of antibacterial drugs within a certain period of time before surgery can achieve this goal. Antibacterial drugs are usually started within 30 minutes to 2 hours before surgery or during the induction of anesthesia, and are used in the operating room rather than in the ward. If the operation time exceeds 3 hours, antibacterial drugs should be added to ensure that the effective concentration of antibacterial drugs in the serum and tissues during the entire operation exceeds MIC90. Under normal circumstances, patients

with type I surgery should not use antibacterial drugs. In special circumstances, antibacterial drugs should be used, and antibacterial drugs should be stopped within 24 hours after surgery. The proportion of patients with type I should not exceed 30%; patients with type II should be discontinued within 48 hours after surgery; and for type III patients should be stopped within 3–7 days. Studies have shown that there is premature medication during the operation period (average treatment time before surgery is 36.3 hours). Patients receive antibacterial treatment in the ward first, and the postoperative drug withdrawal time is later (mean medication time is 36.3 hours). The drug withdrawal time after the operation was 4 days, and the preventive effect of antibacterial drugs after the operation was not as good as that of the preoperative drugs. The use of antibacterial drugs in advance or long-term use of antibacterial drugs after surgery is not beneficial for preventing hospital infections, but it will prolong the overall medication time of patients and increase medical expenses. It may even lead to the production of drug-resistant bacteria and increase the risk of nosocomial infection [28].

The results of the study showed that the pathogenic microorganisms in various parts of hospital infection in surgical patients were mainly Gram-negative bacteria, accounting for 59.0%. The second is Gram-positive bacteria, accounting for 27.7%, and fungi accounting for 8.2%, similar to previous research results. The main reason for this result is the extensive use of broad-spectrum antibacterial drugs, which inhibit the growth of most bacteria, but enable the fungi to multiply selectively. This has brought a lot of troubles in clinical practice and threatened the life safety of patients. According to the number of isolated strains, they are *Escherichia coli*, *Pseudomonas aeruginosa*, Acinetobacter baumannii, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Staphylococcus* hemolyticus, *Candida* albicans, *Enterococcus faecalis*, *Staphylococcus* epidermidis, and Intestinal *Bacillus*. The abovementioned pathogens are the more common pathogens in hospital infections, and the pathogens are selectively grown after extensive use of cephalosporin antibacterial drugs. The microbiological examination rate of hospitalized patients receiving antimicrobial treatment is not less than 30%. Conditional medical institutions should rationally select antibacterial drugs based on the results of clinical microbial specimen testing. Research shows that the microbiological examination rate of surgical patients is 38.2%, which meets the requirements. However, most surgical patients in this hospital have used antibacterial drugs before being submitted for microbiological examination, and there is an empirical phenomenon of drug use. Therefore, hospital infection professionals should increase the training of hospital infection knowledge, and correct some doctors' empirical choice of antibacterial drugs or direct use of antibacterial drugs with a wider antibacterial spectrum. According to the susceptibility test, antibiotics should be correctly selected to treat patients with nosocomial infections, so that doctors can establish the correct medication viewpoints and control the emergence of drug-resistant strains.

6. Conclusion

There are significant differences in the incidence of infections caused by different pathogens in nephrology patients. The study of the article showed that the infection rate of *Klebsiella pneumoniae* reached 33.98%, and the infection rate of *Staphylococcus aureus* was the lowest at 6.80%. Therefore, in the Department of Nephrology, we must focus on strengthening the prevention of *Klebsiella pneumoniae* infection. There are many risk factors for infection in patients with kidney disease; among them, the older the patient, the greater the surgical invasiveness; the fewer the types of antibacterial drugs, the longer the hospital stay, and the use of glucocorticoids, the higher the risk of infection. According to the risk factors of infection, targeted prevention and clinical treatment of various infection factors are carried out to reduce the probability of infection caused by risk factors. Comprehensive monitoring and retrospective research are the foundation, which is the stage of exploring and improving hospital infection professionals to supplement experience. However, retrospective examinations are prone to deviations. Often due to the incomplete records of the original medical records, many cases of infection are undetectable, and missed diagnosis is inevitable. The survey results may not reflect the actual level of hospital infections. With the in-depth advancement of nosocomial infection control, nosocomial infection outbreak surveillance in the top three hospitals has achieved a qualitative leap from comprehensive surveillance investigations to targeted investigations in surveillance. This is essential for timely detection of nosocomial infection epidemics and outbreaks. Through the comprehensive monitoring of nosocomial infections and the continuous development of research methods, the education of the importance of reporting nosocomial infections to the majority of medical staff will be strengthened. Focus on strengthening the complete process of reporting, statistics, and analysis of infectious diseases, hospital infections, and other infections, establishing a hospital medical quality management system, real-time monitoring of hospital infections, and improving work efficiency. The ward during hospitalization should also be ventilated in time to maintain air circulation, and timely cleaning and disinfection can also minimize the spread of bacterial pathogens. At the same time, when formulating a clinical treatment plan, invasive operation treatment measures should be reduced, treatment equipment should be strictly managed to avoid bacterial accumulation, and medical staff should try to ensure aseptic operation to reduce the risk of infection of patients from the source. In addition, measures such as linking the underreporting of nosocomial infections with economic benefits have been implemented in clinics and groups. Therefore, the implementation of nosocomial infection monitoring must be based on a comprehensive and comprehensive nosocomial infection, and the data obtained will be more true and reliable. Studies have shown that about 5% of the infections of surgical patients appear after discharge from the hospital. This study only conducted research on all surgical patients during hospitalization, and did not follow-up on the discharged surgical patients. There is a certain information bias. Therefore, in future studies, discharged patients should be followed up to further understand the precise situation of the incidence of nosocomial infection. This study only involves a comprehensive tertiary hospital, and the results of the study may be different from other studies. The number of hospitals should be expanded, and the sample size should be increased to further understand the effect and feasibility of nosocomial infection surveillance of surgical patients.

Data Availability

No data were used to support this study.

Conflicts of Interest

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

Authors' Contributions

Jiali Zheng and Jiuying Fei are co-first-authors.

References

[1] F. Aktar, R. Tekin, A. Güneş et al., "Determining the independent risk factors and mortality rate of nosocomial infections in pediatric patients," *BioMed Research International*, vol. 2016, Article ID 7240864, 6 pages, 2016.

[2] B. Zhang, X. L. Wu, and R. Li, "A meta-analysis on evaluation of nosocomial infections amongst patients in a tertiary care hospital," *Journal of Healthcare Engineering*, vol. 2021, Article ID 4386423, 6 pages, 2021.

[3] M. A. Osman, M. Alrukhaimi, G. E. Ashuntantang et al., "Global nephrology workforce: gaps and opportunities toward a sustainable kidney care system," *Kidney International Supplements*, vol. 8, no. 2, pp. 52–63, 2018.

[4] S. Saint, "Can intersectional innovations reduce hospital infection?" *Journal of Hospital Infection*, vol. 95, no. 2, pp. 129–134, 2017.

[5] S. Kim, Yeong, and Jihoon, "An agent-based analysis study on maternity clinical space vulnerable to hospital infection," *Journal of Digital Design*, vol. 17, no. 2, pp. 104–114, 2017.

[6] S. Hao, F. Tian, Z. Liang et al., "Sensor array optimization of electronic nose for detection of bacteria in wound infection," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 9, pp. 7350–7358, 2017.

[7] A. Abedallah Zaid and A. Ali, "Alwan, multi-objective chaotic butterfly optimization with deep neural network based sustainable healthcare management systems," *American Journal of Business and Operations Research*, vol. 4, no. 2, pp. 39–48, 2021.

[8] S. David, J. Andrew, K. M. Sagayam, and A. A. Elngar, "Augmenting security for electronic patient health record (ePHR) monitoring system using cryptographic key management schemes," *Fusion: Practice and Applications*, vol. 5, no. 2, pp. 42–52, 2021.

[9] M. K. Kiruthika, S. Gayathri, R. Hemalatha et al., "Design and development of mobile healthcare application for "ayurvedic" based clinical documents," *Journal of Cognitive Human-Computer Interaction*, vol. 1, no. 1, pp. 18–27, 2021.

[10] H. Wei and N. Kehtarnavaz, "Semi-supervised faster rcnn-based person detection and load classification for far field

video surveillance," *Machine Learning and Knowledge Extraction*, vol. 1, no. 3, pp. 756–767, 2019.

[11] G. Chen, Y. Lu, B. Li, K. Tan, and T. Moscibroda, "Mp-rdma: enabling rdma with multi-path transport in datacenters," *IEEE/ACM Transactions on Networking*, vol. 27, no. 6, pp. 1–16, 2019.

[12] X. H. Zou, Y. P. Zhu, and G. Q. Ren, "Significance of bacteria detection with filter paper method on diagnosis of diabetic foot wound infection," *Zhonghua Shaoshang Zazhi*, vol. 33, no. 2, pp. 83–88, 2017.

[13] J. Song, A. Tark, and E. L. Larson, "The relationship between pocket hematoma and risk of wound infection among patients with a cardiovascular implantable electronic device: an integrative review," *Heart & Lung*, vol. 49, no. 1, pp. 92–98, 2020.

[14] A. Avkb, A. Merb, and A. Oqg, "Can natural language processing provide accurate, automated reporting of wound infection requiring reoperation after lumbar discectomy? - ScienceDirect," *The Spine Journal*, vol. 20, no. 10, pp. 1602–1609, 2020.

[15] A. Hb, B. Aj, and C. Hs, "Risk of type III secretion systems in burn patients with *Pseudomonas aeruginosa* wound infection: a systematic review and meta-analysis," *Burns*, vol. 47, no. 3, pp. 538–544, 2021.

[16] H. Zhang, W. Qu, M. Nazzal, and J. Ortiz, "Burn patients with history of kidney transplant experience increased incidence of wound infection," *Burns*, vol. 46, no. 3, pp. 609–615, 2020.

[17] M. Ross, S. Iyer, K. R. Gundle, and D. A. Ross, "P25. Association of preoperative hemoglobin A1c and body mass index with wound infection rate in spinal surgery," *The Spine Journal*, vol. 20, no. 9, pp. S158–S159, 2020.

[18] A. Davies, L. Teare, and M. Shah, "Consensus demonstrates four indicators needed to standardize burn wound infection reporting across trials in a single-country study (ICon-B study)," *Journal of Hospital Infection*, vol. 106, no. 2, pp. 217–225, 2020.

[19] T. Nakamura, Y. Takayama, T. Sato, and M. Watanabe, "Risk factors for wound infection after laparoscopic surgery for colon cancer," *Surgical Laparoscopy Endoscopy & Percutaneous Techniques*, vol. 30, no. 1, pp. 45–48, 2020.

[20] K. K. Mcmackin, G. Ghobrial, M. A. Fadoul, and J. V. Lombardi, "Herpes wound infection after femoral endarterectomy," *Journal of Vascular Surgery Cases, Innovations and Techniques*, vol. 6, no. 2, pp. 185–187, 2020.

[21] A. Yang, M. Yassin, and T. Phan, "Vibrio mimicus wound infection in a burn patient," *Radiology Case Reports*, vol. 16, no. 6, pp. 1348–1351, 2021.

[22] F. I. Jahan, F. Begum, and F. Islam, "Risk factors for wound infection following caesarean section: a cross sectional cohort study from sir salimullah medical college & mitford hospital in dhaka, Bangladesh," *Open Journal of Obstetrics and Gynecology*, vol. 09, no. 6, pp. 904–913, 2019.

[23] F. Biancari and S. Giordano, "Glycated hemoglobin and the risk of sternal wound infection after adult cardiac surgery: a systematic review and meta-analysis," *Seminars in Thoracic and Cardiovascular Surgery*, vol. 31, no. 3, pp. 465–467, 2019.

[24] S. Ohira, H. Miyata, S. Yamazaki et al., "Deep sternal wound infection after bilateral internal thoracic artery grafting: insights from a Japanese national database," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 157, no. 1, pp. 166–173, 2019.

[25] M. Tanriverdi, "A systematic review of privacy preserving healthcare data sharing on blockchain," *Journal of Cybersecurity and Information Management*, vol. 4, no. 2, pp. 31–37, 2020.

[26] G. Pomponio, S. Tedesco, A. Peghetti et al., "Improving the quality of clinical research on chronic wound infection treatment: expert-based recommendations," *Journal of Wound Care*, vol. 28, no. Sup1, pp. S26–S31, 2019.

[27] A. Gs, B. Mr, and C. Ss, "Prevention of deep sternal wound infection in cardiac surgery – reply to Vos et al," *Journal of Hospital Infection*, vol. 102, no. 3, pp. 295-296, 2019.

[28] E. Rba, A. Ab, and B. Eo, "Hyperglycemia as a risk factor for postoperative early wound infection after bicondylar tibial plateau fractures: determining a predictive model based on four methods," *Injury*, vol. 50, no. 11, pp. 2097–2102, 2019.

Hindawi

Retraction

Retracted: Correlation Meta-Analysis of the Efficacy of Inhaled Corticosteroids in Children with Asthma Based on Smart Medical Health

Journal of Healthcare Engineering

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

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

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

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

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

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

References

[1] Y. Qin, J. Wang, J. Qin et al., "Correlation Meta-Analysis of the Efficacy of Inhaled Corticosteroids in Children with Asthma Based on Smart Medical Health," *Journal of Healthcare Engineering*, vol. 2022, Article ID 6220774, 12 pages, 2022.

Research Article

Correlation Meta-Analysis of the Efficacy of Inhaled Corticosteroids in Children with Asthma Based on Smart Medical Health

Yu Qin,[1] Jing Wang,[2] Jingmin Qin,[3] Ning Yang,[1] Sha Li,[1] Lijia Xu,[1] and Yanjun Han [1]

[1]*Pediatrics, Xingtai People's Hospital, Zhanjiang 054031, Hebei, China*
[2]*Endocrine Department, Xingtai People's Hospital, Zhanjiang 054031, Hebei, China*
[3]*Neonatology, Handan Second Hospital, Handan 056001, Hebei, China*

Correspondence should be addressed to Yanjun Han; 201904012228@stu.zjsru.edu.cn

Children are the reserve force for national construction, and children's health is now being paid more and more attention. In today's childhood asthma treatment, many are based on the treatment of common cough, and there is no strict division with asthma. Importantly, the efficacy of inhaled corticosteroids in the treatment of childhood asthma has been extensively studied in the medical community. However, there is no clear explanation for its specific efficacy and role. Research on childhood asthma has always been a key topic in medicine. This article aims to conduct a meta-analysis of the correlation between inhaled corticosteroids and asthma in children. This article proposes a more scientific literature screening strategy, which can achieve a higher accuracy rate for meta-analysis. It combines the intelligent medical health system in the analysis of children's asthma to assist in the analysis of curative effects. Such analysis also provides reference significance for related research. The experimental results indicate that inhaled corticosteroids have a good effect on childhood asthma. As the concentration of inhaled glucocorticoid increases, the effect will be better. Specifically, if the concentration is increased by 10%, the effect will be about 15% better, but there are corresponding side effects.

1. Introduction

With the advancement of computer information technology, the medical field has increasingly computerized and networked various management forms to reduce costs, facilitate management, and improve the effectiveness of treatment. Early foreign studies have found that the remote management system using telephone, mobile phone text messages, and the Internet as information technology can shorten the distance between the doctor and the patient and help patients adjust their diet and exercise patterns in time. It strengthens blood glucose monitoring and improves the level of disease control. Asthma is a chronic respiratory disease with a high clinical incidence in pediatrics. Although most children can be effectively controlled after effective treatment, there are still some cases of disease that are prolonged and difficult to heal. How to improve the therapeutic effect of asthma is a hot spot in clinical research. Inhaled corticosteroids are the first-line drugs for the treatment of asthma and have definite effects on children with persistent asthma of varying severity. It has been listed as a recommended drug in the global asthma prevention and treatment strategy since 2006.

However, the specific principles and corresponding side effects of inhaled corticosteroids on childhood asthma are not particularly clear. For related studies, more researches are conducted on asthma cases, and there is no comprehensive correlation analysis. Especially with the current technological development of smart medical health, it is very necessary to carry out a correlation analysis of the efficacy of inhaled corticosteroids in children's asthma based on smart medical health.

Therefore, there are two main innovations in this article: (1) In the analysis, it focuses on combining modern smart health medical technology. Based on this, it introduces the application scenarios of smart medical treatment and related technologies. (2) The literature screening strategy in the meta-analysis has been greatly optimized, and some influential and useless factors have been eliminated so that the results of the meta-analysis will be more realistic.

2. Related Work

There are different studies on the efficacy of inhaled corticosteroids. Parameswaran K and others believe that there is no consensus on the method of comparing the clinical efficacy of different inhaled corticosteroids. Comparisons need to be made in terms of relative efficacy, and the study should include comparisons of two or three doses. Drug deposition studies and mathematical models of drug pharmacokinetics in the airway can provide supplementary information for clinical drug relative efficacy studies. When estimating the clinical and systemic effects, the fine particle dose and the emitted dose should be considered separately instead of the nominal dose. When comparing with an innovative drug (same compound and same equipment) to evaluate the second entry (generic) drug, if the generic drug meets drug equivalence, it may be appropriate to consider accepting the generic drug as bioequivalence [1]. Qian et al. have studied the side effects of long-term use of inhaled corticosteroids. They believe that these side effects include impaired growth in children, osteoporosis, fractures, glaucoma, cataracts, and thinning of the skin. So they will review the current recommendations on the use of ICS in asthma management to highlight the treatment strategy that strikes the best balance between safety and effectiveness [2]. Hihoriya et al. conducted a randomized and prospective study to determine the efficacy of montelukast as a $\beta 2$ agonist and inhaled corticosteroids in patients with moderate persistent asthma. Experimental results show that montelukast treatment can better improve the clinical symptoms and PFT of patients with moderate persistent asthma [3]. Ema et al. believe that the treatment of the "yellow zone" or worsening of childhood asthma is still controversial. Their goal is to review existing data and consider the effect of increasing the dose of ICS and potential alternatives to this common practice [4]. Hong et al. believe that although nebulized corticosteroids (NebCS) are a key treatment option for young children with asthma or viral wheezing (VIW), there is no uniform recommendation for their best use. The purpose of their study was to clarify the role of NebCSs in the management of acute exacerbations of asthma, asthma maintenance treatment, and VIW treatment in children 5 years of age or younger. The results prove that NebCSs are effective and well-tolerated in the treatment of acute and chronic asthma in patients aged 5 years or younger [5]. Maio et al. believe that dual-agent chemotherapy as the first-line treatment for advanced nonsmall cell lung cancer (NSCLC) is more effective than single-agent chemotherapy. As a second-line treatment, several randomized trials comparing single-drug and dual-drug chemotherapy have been conducted, but each trial is insufficient to detect potential related survival differences. They proved through experiments that dual-agent chemotherapy as a second-line treatment for advanced NSCLC significantly improved the remission rate and progression-free survival. However, it is more toxic than single drugs and cannot improve overall survival [6]. There are also more studies on meta-analysis. Tambini and Roberto conducted a meta-analysis of 5 randomized controlled trials (RCTs) involving 339 patients with acute infectious mononucleosis (IM) treated with acyclovir (ACV). They believe that clinical data does not support the use of ACV for the treatment of acute IM, although the drug has good virological activity [7]. Chu et al. conducted a meta-analysis on suicide and related diseases. Their research conclusion supports the theory of interpersonal relationships: the interaction between frustrated belonging and the perceived burden is significantly related to suicidal ideation. The interaction between frustrated belonging, perceived burden, and suicidal ability is significantly related to more previous suicide attempts. However, the impact of these interactions is moderate. Alternative configurations of theoretical variables are equally useful for predicting suicide risk and theoretically consistent paths. He summarized the limitations and suggestions of interpersonal relationship theory as a framework for understanding the spectrum of suicide [8]. However, there is an obvious deficiency in these theories. That is, modern intelligent medical technology has not been combined and modern medical technology has not been applied, which is not conducive to the combination of technological development and modern medical care. And their research is more on the treatment of asthma, without distinguishing the special group of children. However, this paper pays more attention to the group of children. When conducting meta-analysis, an important point in the search strategy is to limit children. Such strict restrictions on children as research subjects can clearly study the specific effects of inhaled glucocorticoids on children.

The first part of the article introduces the background, significance, and related work of the research on the efficacy of inhaled glucocorticoids in children with asthma and proposes the innovations of this paper. The meta-analysis of childhood asthma and inhaled corticosteroids is proposed, and the intelligent medical treatment based on the Internet of Things is analyzed. The fourth part proposes a meta-analysis of the efficacy correlation and proposes a more scientific literature screening strategy, which can achieve better results. High meta-analysis accuracy: Combined with the intelligent medical health system, it is used to assist the analysis of the curative effect in the analysis of children's asthma, and then the results of the experimental part are analyzed and summarized.

3. Introduction of Smart Medical and Health Related Technologies

3.1. Childhood Asthma and Inhaled Corticosteroids.
Childhood asthma seriously harms children's health worldwide. For many years, it has become the most common public health problem in childhood [9]. The incidence and mortality of bronchial asthma in children are on the rise worldwide. Through in-depth investigation and clinical research analysis among clinical workers in many countries

and regions, it has been found that the prevalence of bronchial asthma in children has increased significantly in the past 10 years. The incidence and prevalence of asthma in the world have also increased substantially [10]. In recent years, the prevalence of bronchial asthma in children in China has gradually increased. There are currently about 20 million asthma patients in China. In 1996, the domestic pediatric bronchial asthma investigation team conducted a survey of normal and ill children aged 6–12 years in various provinces, regions, and counties in China. The results found that the incidence of bronchial asthma in children in China has increased significantly compared with the previous period. By the end of 2010, the results of the survey of childhood asthma (6–12) in existing cities in China showed that the incidence of asthma was 0.25%–4.63%, with an average of 1.97%. The results of the reinvestigation in 2006 were 0.35%–4.38%, an average of 2.34%, and an average increase of 74.84% over the past 10 years. Among them, Shanghai is 5.24%, currently ranking first in the country [11, 12]. As humans continue to transform and pollute the Earth, a series of changes will occur in the current atmospheric environment, biological species and ecology, and dietary structure. Therefore, allergic symptoms and bronchial asthma will gradually show an upward trend. Asthma is mainly composed of a variety of inflammatory cells and cellular components that participate in the onset and pathological evolution of asthma. The main clinical manifestations of patients have increased airway responsiveness and recurrent wheezing, shortness of breath, chest tightness, and (or) coughing. At present, inhaled corticosteroids are the most effective anti-inflammatory drugs for the first-line treatment of asthma. It mainly reduces the chronic inflammation of the airway by reducing the activation of T lymphocytes and inhibiting the production of cell-related inflammatory factors [13]. Helper T lymphocytes also play an important role in immunity and regulation in chronic airway inflammation in children with asthma [14].

With the in-depth clinical understanding of the pathogenesis of asthma, people have gradually gained new understanding and research on the mechanism of action of glucocorticoids [15]. Through long-term inhalation of glucocorticoids to understand the pathological and related physiological changes of children, it is helpful to clarify the influence of glucocorticoids on the intrinsic immune function of children. It can provide a reliable basis for the Formulation of glucocorticoid medication regimens in clinical research and the in-depth study of the specific mechanism of glucocorticoids [16]. A large amount of literature analysis pointed out that IgE, TNF-α, IL-8, IL-17,

LTB4, and other inflammatory factors played an important role in the occurrence, development, and evolution of asthma. Therefore, glucocorticoid anti-inflammatory therapy is an important part of the treatment of asthma. In reality, we mainly use inhaled corticosteroids to treat asthma in clinical practice, which has a definite effect [17]. Therefore, studying the related pathogenesis and new treatments of bronchial asthma has become a global health problem to be solved. Therefore, we should pay enough attention and work hard to solve it from now on. Active and reasonable detection of children with bronchial asthma can prompt the survival and comfort of the children, improve the symptoms and duration of asthma attacks, and reduce the economic burden of children with children [18].

3.2. Meta-Analysis. There are two main schools of statistics: the classical frequency statistics school and the Bayesian school scholars have contributed to the development of network meta-analysis (NMA). And some scholars have compared the application of the two faction theories in meta-analysis [19]. The frequency school's NMA method is based on the assumption that the effect scale obeys normal distribution. Bayesian statistics takes the target parameter θ as a random variable with a priori distribution. After obtaining the sample information, it obtains comprehensive information about the posterior distribution $\pi(\theta|x)$ of θ through simulation iteration so as to obtain the Bayesian estimation of the parameter θ. In recent years, it is based on the Bayesian NMA viewpoint put forward by Higgins and Whitehead. Many scholars have gradually developed multiple models such as simple Bayesian NMA, NMA under the framework of generalized linear models, and multilevel hierarchical Bayesian NMA, collectively referred to as "Bayesian NMA model" [20]. The following is an introduction to the NMA model algorithm based on Bayesian model averaging.

The symbol β represents the parameter vector to be estimated, and N_k represents the k-th model in the better model space. Let A denote the observed data sample, the posterior density distribution of β can be written in the following form:

$$P(\beta|A) = \sum_{K-1}^{K} P(\beta|A, N_K)P(N_K|A).\tag{1}$$

In formula (1), A represents a set of i samples. According to formula (1), the estimated value and variance of the parameter vector β to be solved are as follows:

$$E(\beta|A) = \int_{-\infty}^{+\infty} \beta \left[\sum_{K-1}^{K} P(\beta|A, N_K)P(N_K|A) \right] d\beta,\tag{2}$$

$$Var(\beta|A) = \int_{-\infty}^{+\infty} [\beta - E(\beta|A)]^2 \left[\sum_{K-1}^{K} P(\beta|A, N_K)P(N_K|A) \right] d\beta,\tag{3}$$

or abbreviated as follows:

$$E(\beta|A) = \widehat{\beta}P(N_K|A), \qquad (4)$$

$$\text{Var}(\beta|A) = V1 + V2. \qquad (5)$$

In formula (5), $V1$ is the error between models, and $V2$ is the error within the models, expressed as follows:

$$V1 = \sum_{K-1}^{K} P(N_K|A)\left(\widehat{\beta} - \sum_{K-1}^{K} P(N_K|A)\widehat{\beta}\right)^2, \qquad (6)$$

$$V2 = \sum_{K-1}^{K} P(N_K|A)Var^2(\beta|A, N_K). \qquad (7)$$

It can be seen from the introduction that determining the posterior probability of the model is the key to the model. The parameter conditions make the model optimal. Then the posterior probability is as follows:

$$P(\Delta|N_K, A) = \int P(\Delta|\beta_K, N_K, A)P(\beta_K|N_K, A)d_\beta. \qquad (8)$$

In formula (8), $(1, 2, ..., k)$ is the unique regression parameter vector of the better model N_k. Use maximum likelihood estimation to approximate as follows:

$$P(\Delta|N_K, A) \approx P(\Delta|N_K, \widehat{\beta}_K, A). \qquad (9)$$

This estimate can be used as a very good approximation when it is considered uncertain about the model in subsequent proofs, and it is used by subsequent scholars [21, 22].

Write the model's posterior probability estimate (BIC) as a Bayesian Formula, as shown in the following formula:

$$P(N_K|A) = \frac{P(A|N_K)P(N_K)}{\sum_{J=1}^{K} P(A|N_K)P(N_K)}, \qquad (10)$$

In formula (10), $P(N_k)$ is the prior probability of model N_k, so formula (10) can be written as follows:

$$P(N_K|A) = \frac{P(A|N_K)}{\sum_{J=1}^{K} P(A|N_K)}. \qquad (11)$$

In formula (11), $P(A|N_k)$ can be written in the form of a marginal likelihood function, that is,

$$P(A|N_K) = \int P(A|\beta_K, N_K)P(\beta_K|N_K)d\beta_k. \qquad (12)$$

The estimated value of formula (12) is obtained by the Laplace approximation method as shown in the following formula [23]:

$$\log(P(A|N_K)) = \log(P(A|\beta, N_K)) - \left(\frac{d_k}{2}\right)\log n + o(1). \qquad (13)$$

Kd in formula (13) represents the number of parameters to be estimated in the k-th model N_k, and n represents the number of samples. The Bayesian information criterion is as follows:

$$BIC_K = -2\log(P(A|N_K, \beta_K)) + d_k\log n. \qquad (14)$$

It is further transformed into the following formula:

$$-\frac{BIC_K}{2} = \log(P(A|N_K, \beta_K)) - \left(\frac{d_k}{2}\right)\log n. \qquad (15)$$

Therefore, the posterior probability expression of the BIC model is as shown in thefollowing formula:

$$P(N_K|A) = \frac{\exp(-BIC_K/2)}{\sum_{J=1}^{K} \exp(-BIC_K/2)}. \qquad (16)$$

So far, the NMA model algorithm based on Bayesian model averaging has been introduced. The uncertainty of models and parameters and the comprehensive and effective use of information are the main factors that affect the results of meta-analysis. This paper uses the Bayesian Model Average (BMA) method to model and optimize the meta-analysis and integrates the information of the alternative models and variables to control the uncertainty of the model and effectively utilize the rich key information.

3.3. Smart Healthcare. The application of advanced technology in medical testing instruments provides favorable conditions for the effective diagnosis of public pathology, which can fully meet the needs of medical diagnosis [24]. However, the medical testing instruments currently used in the market are usually bulky and expensive. When in use, it needs to be in close contact with the test object, and the test object cannot move casually. It is generally not suitable for monitoring a large range of high-risk infectious disease patients, and it is not suitable for patient populations that are geographically dispersed but require real-time monitoring [25]. Therefore, through heterogeneous wireless network coordination, a seamless connection between WLAN and mobile communication network is realized in a ubiquitous network environment, which is conducive to the health prevention and control center to grasp the information of the monitored patient at any time and any place. Thus, it provides a strong basis for the adoption of effective diagnosis and treatment plans, which is very feasible and has national strategic application value for preventing and controlling the spread of sudden major infectious diseases.

However, due to the characteristics of China's national conditions, it is difficult to use the network for remote management. The development of medical technology in the new era is rapidly developed in conjunction with the development of other information technologies [26]. In modern medicine, digital, visual information, high-precision sensors, electromagnetic imaging, and other technologies are used to make it more convenient to observe various information of the human body. It makes the diagnosis more accurate, and people can get better medical services. Before the advent of these technical devices, doctors could only obtain the physiological state of patients by means of experience and hearing. This method relies more on the subjective ability of the doctor and has a great risk. At the

same time, the technology is difficult to promote among new doctors [27]. Therefore, the medical diagnosis technology at that time mainly relied on the doctor diagnosed. With the development of electronic information, people have discovered that electromagnetic waves, sound waves, and other information can be used to realize the manifestation of human physiological signals. As a result, the medical diagnosis will no longer rely on the doctor's experience but use high-tech to accurately reflect the physiological changes of the human body [28]. Among them, the application scenarios of smart medical systems can be divided into isolated areas and nonisolated areas, as shown in Figure 1:

As shown in Figure 1, the descriptions of the two scenarios are as follows.

Scenario 1: When a patient is isolated in a designated place, the portable multiphysiological parameter collection terminal will be connected to the wireless AP point in the environment through WiFi. Since the patient is at the isolation point, it is not convenient for medical staff to enter and observe the patient's condition at any time, so it must be linked through the device. However, the network of isolation points needs to be relatively stable to meet the needs of medical staff and patients. For example, in the case of a pneumonia epidemic, the patient is in the isolation room, and all data are tested by wearable devices.

In these two scenarios, a large-scale high-speed mobile network and a small-scale low-power wireless access technology can be organically combined. It realizes the monitoring of infectious epidemics in the two scenarios and ensures a seamless connection in two scenarios. Monitoring personnel can use mobile phones, laptops, and other terminal devices to access the application server of the Health Prevention and Control Center through the Internet, regardless of whether they are working in the quarantine area or when they are outside the quarantine area. It knows the physiological parameters of the monitored object in time so as to make corresponding prevention and control measures.

Among them, its specific architecture and services are shown in Figure 2:

Intelligent medical platforms based on the Internet of Things include cloud computing centers, hospital platforms, home terminals, service centers, etc. As shown in Figure 2, the monitoring system is divided into 4 layers, which are the monitoring terminal, information interaction system, database, and monitoring center. For the monitoring terminal, it is targeted at specific patients. They obtain basic information through their own or hospital-provided testing equipment, or the equipment can take and upload it by themselves. This is very convenient for the patient and for the nurse, the simple and time-consuming measurement operation is avoided, and the patient can solve it by himself. For the second-level information interaction system, the network, data processor, data collector, etc., are integrated to achieve the functions of uploading and data processing. The data collected by patients is simply classified and analyzed through the information interaction system, and simple statistical processing is performed to form a report directly to the doctor and upload the data to the third-tier hospital's

database so that it is more convenient for doctors to make accurate judgments about the patient's overall situation and recent conditions. For the fourth layer, it is aimed at medical staff. They use the testing center to conduct unified testing on patients, and they can respond to the situation as soon as possible. It accurately locates the data of the problem, which is the location where the problem occurs, so that the targeted and timely treatment can be carried out.

4. Meta-Analysis of the Correlation of Curative Effect

4.1. Literature Inclusion Strategy. The inclusion and exclusion of the literature should follow practical principles, and there should be information about the efficacy of inhaled corticosteroids in children. Specifically, the following strategies:

(1) Literature inclusion criteria

① The original documents are observational studies (cohort studies, cross-sectional studies, time series, and case crossover);

② It studied the efficacy of inhalable glucocorticoids;

③ The research object is children, and the diagnostic criteria or methods of asthma are described in the literature;

④ The results of the analysis of the curative effect on asthma are reported in the literature;

⑤ For the repetitive literature, choose the one with the closest topic to this one;

⑥ The languages are limited to Chinese and English.

(2) Document exclusion criteria

① News, letters, summaries, reviews, abstracts, and lectures, etc.;

② It does not describe the diagnostic criteria or methods of childhood asthma;

③ Incomplete and unavailable documents provided by the data;

④ The literature research population includes adults and children, but children's data cannot be extracted separately.

⑤ Publication languages are other than English and Chinese.

(3) Literature search strategy

Documents were screened by computer and searched in the research papers databases of How-Net, Wanfang, and Zhangqiao.

The time is set from 2016 to present, and the papers are roughly screened based on the keywords of the papers, and then the initially selected documents are manually screened. The specific screening steps are shown in Figure 3.

As shown in Figure 3, the first step is to search through keywords and a total of 5,324 papers were screened, and after deleting duplicates, 3491 papers remain. Then the literature was screened according to the principle of exclusion, leaving
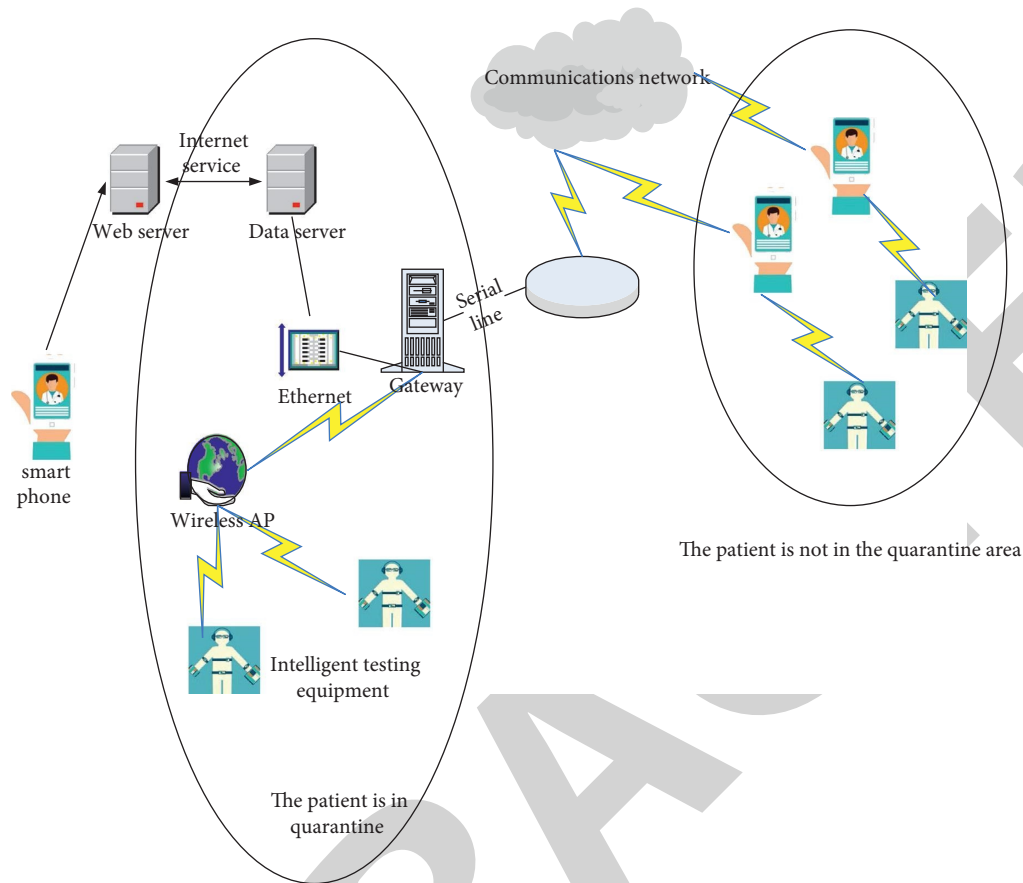
Figure 1: Application scenario diagram of the intelligent medical real-time monitoring system.

189 articles, and after screening the full text, 31 articles remained, and then 31 articles were meta-analyzed.

Among them, the 31 articles after screening are classified, and the country classification is shown in Table 1; the research category is shown in Table 2. Among the 31 documents in Table 1, there are 13 in the United States, 9 in China, 6 in Germany, and 3 in Japan. In Table 2, there are 11 studies for cohort studies, 13 for cross-sectional studies, 7 for case crossover studies, 3 for time series studies, and 19 for inhaled corticosteroids.

4.2. Meta Analysis Results

(1) The quality of the articles after screening is shown in Figure 4.

In Figure 4, based on the degree of suitability between the research content of the literature and this article, the literature quality of the four countries is shown in the figure. It can be found that in the four countries, the quality of the literature is relatively high, with more than 25% of the articles of excellent quality. Among them, because there are only 3 Japanese documents, the excellent rate reached 66.6%. Therefore, the extraction of the literature is quite satisfactory.

(2) Meta result analysis

Thirty-one articles have studied the effect of inhalable corticosteroids on childhood asthma. There is a certain degree of bias in the literature. This paper chooses the random-effects model to merge the results. The results of the meta-analysis are shown in Figure 5. It can be seen in the figure that most of the studies are single studies with a large confidence interval. The diamond is on the vertical line, indicating that there is no difference between the two groups.

(3) Heterogeneity

The heterogeneity analysis of the included 31 articles was carried out, and a funnel chart was drawn for them, as shown in Figure 6.

In Figure 6, the ordinate represents the error. The abscissa is the odds ratio (OR). It can be seen from the figure that most of the points are concentrated at the top, and some points are scattered at the bottom. This shows that most of the samples are large samples, and only a few small samples are scattered at the bottom. For the funnel graph, there are missing corners in the graph, which indicates that the sample has publication bias, and there are two points outside the funnel. For the 95% confidence interval, this means that there is no obvious heterogeneity between the research samples.
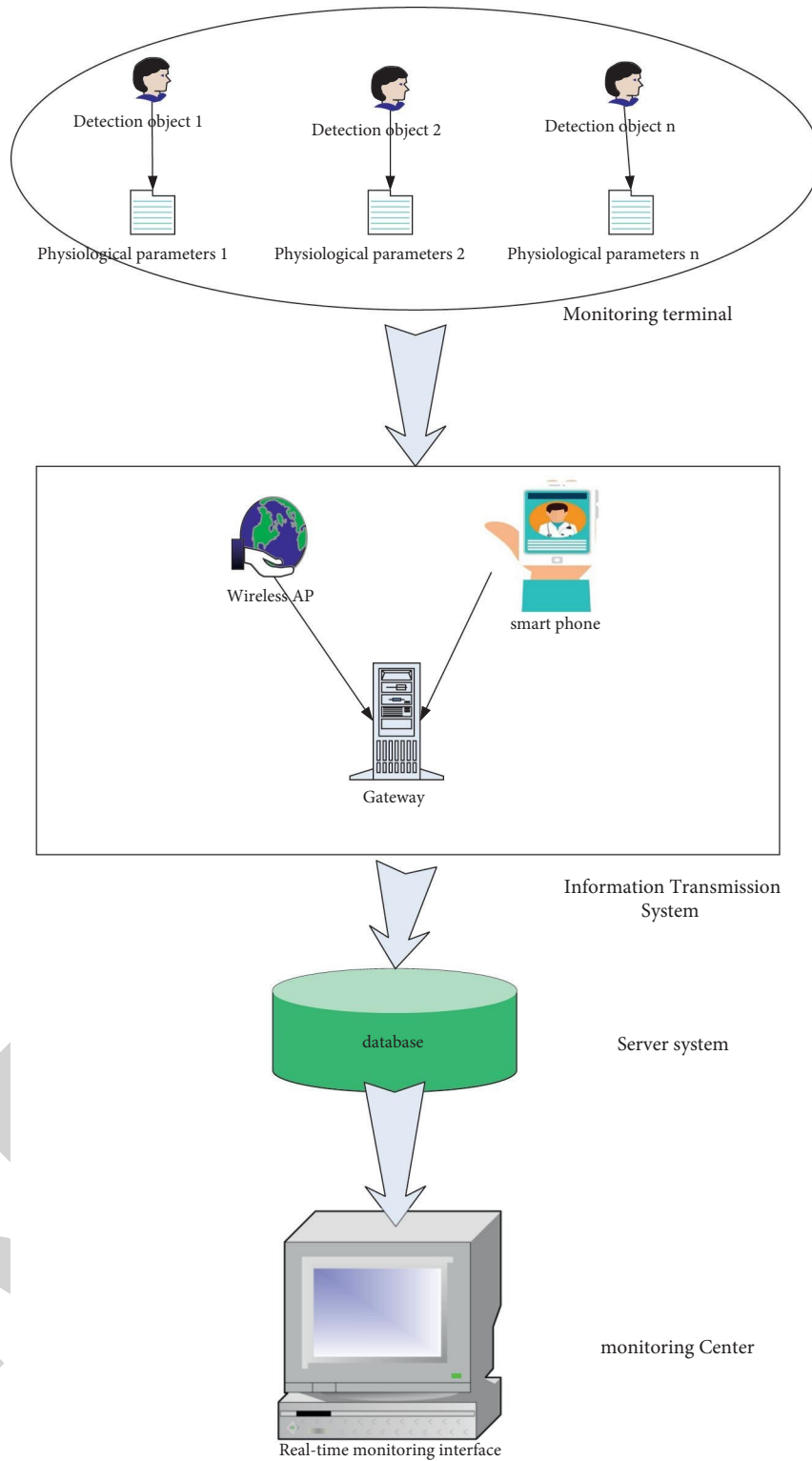
FIGURE 2: The functional block diagram of the intelligent medical real-time monitoring system.
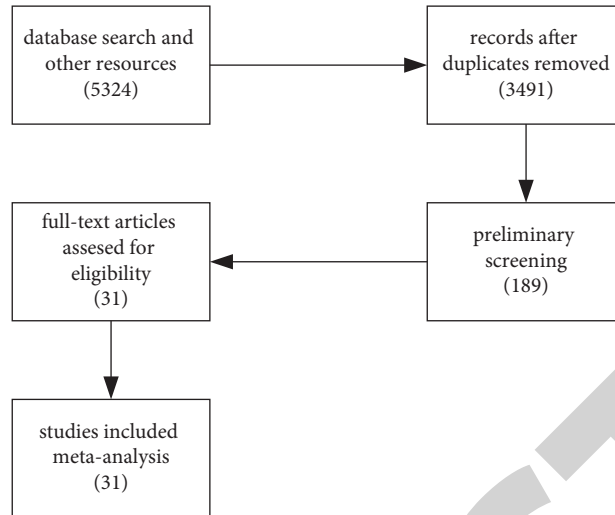
Figure 3: Screening steps diagram.

Table 1: Document country list.

Country	Number of documents	Percentage (%)
America	13	41.9
China	9	29.0
Germany	6	19.4
Japan	3	9.7

Table 2: Literature research category table.

Research category	Number of documents	Percentage (%)
Array research	11	20.8
Cross-sectional study	13	24.5
Case crossover study	7	13.2
Time series research	3	5.7
Inhaled glucocorticoid research	19	35.8
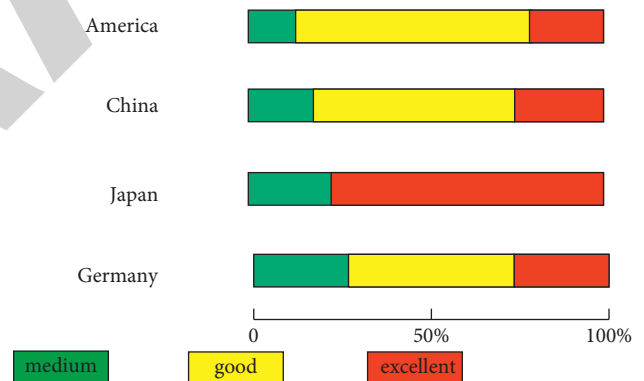
Figure 4: Document quality evaluation chart.

(4) Publication bias

Egger's linear regression method is used for the publication bias test because this method can take into account the existence of the intercept term. As shown in Figure 7, it can be found from the regression graph that the sample points have a downward trend in the graph, and the fitted trend line is also downward. This woman shows that there is a publication bias in the effect of inhalable corticosteroids on asthma. This is the same as the result of the funnel chart in Figure 6, and there is publication bias. It is more obvious that the sample has publication bias because there is a certain bias in the selection of the literature.

(5) Sensitivity analysis

Finally, a sensitivity analysis of the literature is carried out, and we adopt the method of exclusion one by one to carry out the sensitivity analysis. As shown in Figure 8, the results show that no single article can significantly affect the overall heterogeneity. A sensitivity analysis was performed on the results of the overall combined analysis using the one-by-one elimination method. That is, each study was excluded. Meta-analysis was performed again on the remaining studies, and the results showed that the sensitivity analysis results did not change significantly compared with those before the exclusion. The credible interval of each indicator overlaps with the original value to a large extent, suggesting that the results of the meta-analysis of this study are more reliable.

4.3. Efficacy Analysis. The experimental data is an existing data sample in the hospital, which is analyzed on statistical software. To explore the clinical efficacy of NIPPV
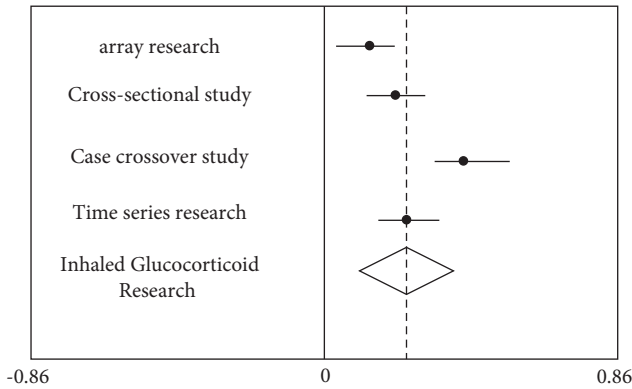
FIGURE 5: The relative risk diagram of the efficacy of inhalable glucocorticoids.

FIGURE 6: OR funnel chart.

FIGURE 7: Regression graph of the effect of inhalable glucocorticoids on childhood asthma.
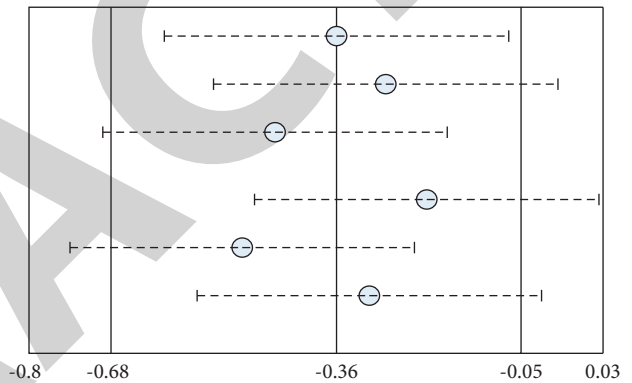
FIGURE 8: Sensitivity analysis graph.

(noninvasive positive pressure ventilation) in the treatment of bronchial asthma. A total of 162 patients with bronchial asthma who were admitted to a hospital from June 2014 to May 2018 were included in the study, and the clinical efficacy and symptom disappearance time of the control group (conventional treatment) and the observation group (inhaled glucocorticoid treatment) were compared.

(1) The efficacy of inhaled glucocorticoids

From Figure 9, we can see that inhaled corticosteroids have a relatively good effect on childhood asthma. When $F \geq 2$, the effect can be at a relatively high value. The highest point can reach 0.98.

(2) Diagnosis and treatment accuracy

As can be seen from Figure 10, the accuracy of inhalable corticosteroids for the diagnosis of asthma is also relatively high. The accuracy of diagnosis can reach more than 90% at different concentrations, and with the increase of time, the accuracy of diagnosis and treatment will become higher and higher.

5. Discussion

Childhood asthma is a source of greater harm to children's health and may cause irreversible damage to lung damage. At present, there are also some medicines for treating children's asthma, and there are many treatment methods. However, there are still relatively few treatments for radical cures or no side effects. After all, children are in a poor physique. Corresponding body functions have not yet been fully perfected. Too much medicine or adult medicine is not reasonable. This article refers to the research of other scholars and adopts the method of meta-analysis. The 31 documents included in the meta-analysis are all observational studies. Among them, there are 11 cohort studies, which are a type of research design with better etiology, and the overall sample size of the included literature is relatively large. The cohort studies and case crossover studies included in this article use the NOS score to assess the risk of bias. Of the 18 articles, 12 have a score greater than or equal to 8. The included cross-sectional studies were evaluated using the AHQR scale, and the coincidence rates of 9 items were all high. The quality of the literature entered into this analysis is relatively high, and there is a dose-effect relationship between acute and chronic effects. There is a certain degree of clinical heterogeneity in the literature included in this article, such as different regions and different age groups. Therefore, a further subgroup analysis of inhalable glucocorticoids is still not possible to completely eliminate the heterogeneity in the literature. At the same time, the chronic effect literature has obvious publication bias. Combined with the GRADE evidence quality
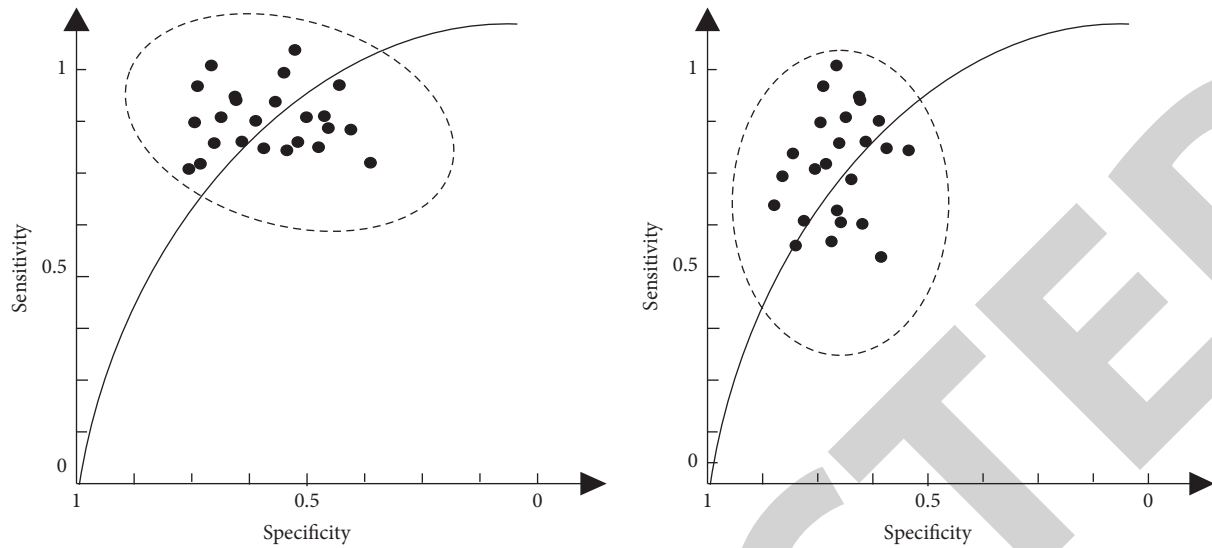
Figure 9: Fibroscan diagnosis $F \geq 2$.
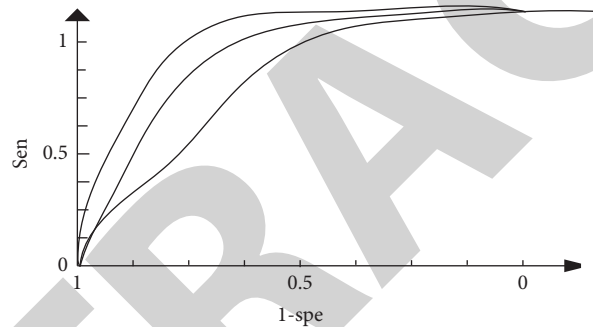
Figure 10: Three SROC curve of sCD14-ST diagnosis accuracy of asthma.

evaluation tool, the quality of evidence for the efficacy of inhaled glucocorticoids on childhood asthma is relatively high. However, the principle of its side effects is still not clear enough, and some related studies believe that it will affect the height development of children. Therefore, the efficacy of inhaled glucocorticoids remains to be analyzed.

However, the limitations of this study are as follows: (1) The quality of the included RCTs (Used to evaluate the possibility of bias) is not the same. Although they all have the quality that can be included, high-quality, high-level RCTs are more important in future studies. In future research, the grouping information and methods, as well as the situation of loss to follow-up and withdrawal, should be described in detail. (2) There must be heterogeneity between different studies, which will have a certain impact on the results of meta-analysis. These may be caused by different basic conditions, such as age, gender, and disease course. Moreover, the different types and doses of oral drugs in the control group may also affect the results. (3) The existence of publication bias is always a problem, and differences in research results and sample size can affect publication bias. This may be related to negative results. That is, articles that do not have any significant differences in results are more difficult to publish. (4) The research on smart medical technology has

not fully targeted children's asthma. The literature is not deep enough for its research, which means that the article has not fully used the smart medical technology of children's asthma.

6. Conclusions

This study systematically evaluated the efficacy of inhalable glucocorticoids on children with asthma and preliminarily summarized its laws in the various research literature, with a meta-analysis as a result. When the two are not affected by the prevalence, the value range is between (0, 1), and the closer to 1, the higher the diagnostic accuracy of the experiment to be evaluated. In addition, as the pathological stage increases, the sensitivity and specificity values gradually increase. This indicates that the efficacy of inhalable glucocorticoids increases with the increase in concentration, and the misdiagnosis rate gradually decreases. The article also emphatically mentioned the application of smart medicine and analyzed its structure and application. It is hoped that smart medical technology can be applied to the case study of childhood asthma in the follow-up research, and combined with the technology of the system platform, to deeply explore the curative effect and influence of inhaled corticosteroids on childhood asthma.

Data Availability

No data were used to support this study.

Conflicts of Interest

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

Acknowledgments

References

[1] K. Parameswaran, R. Leigh, P. M. O'Byrne et al., "Clinical models to compare the safety and efficacy of inhaled corticosteroids in patients with asthma," *Canadian Respiratory Journal*, vol. 10, no. 1, pp. 27–34, 2016.

[2] Y. Qian, X.-O. He, and A. D'Urzo, "A review on the safety and efficacy of inhaled corticosteroids in the management of asthma," *Pulmonary Therapy*, vol. 3, no. 1, pp. 1–18, 2017.

[3] N. Hihoriya, P. Shelat, J. Kagathara, and S. Kumbar, "A prospective, open labeled, comparative study to assess the efficacy of montelukast as add on to β2-agonist and inhaled corticosteroid in patients of moderate persistent asthma," *International Journal of Basic & Clinical Pharmacology*, vol. 1, no. 2, pp. 91–101, 2017.

[4] A. Ema, B. Sjs, and A. Abb, "What is the role of increasing inhaled corticosteroid therapy in worsening asthma in children?" *Journal of Allergy and Clinical Immunology: In Practice*, vol. 7, no. 3, pp. 842–847, 2019.

[5] J. G. Hong, G. Wandalsen, K. R. Murphy et al., "Nebulized inhaled corticosteroids in asthma treatment in children ≤5 Years of age: a systematic review and global expert analysis," *Journal of Allergy and Clinical Immunology: In Practice*, vol. 8, no. 6, pp. 1815–1827, 2020.

[6] M. D. Maio, P. Chiodini, V. Georgoulias et al., "Meta-analysis of single-agent chemotherapy compared with combination chemotherapy as second-line treatment of advanced non-small-cell lung cancer," *Journal of Clinical Oncology*, vol. 27, no. 11, pp. 1836–1843, 2016.

[7] T. Tambini and T. Roberto, "Acyclovir for treatment of infectious mononucleosis: a meta-analysis," *Scandinavian Journal of Infectious Diseases*, vol. 31, no. 6, pp. 543–547, 2017.

[8] C. Chu, J. M. Buchman-Schmitt, I. H. Stanley et al., "The interpersonal theory of suicide: a systematic review and meta-analysis of a decade of cross-national research," *Psychological Bulletin*, vol. 143, no. 12, pp. 1313–1345, 2017.

[9] A. Beigelman and L. B. Bacharier, "Management of preschool children with recurrent wheezing: lessons from the NHLBI's asthma research networks," *Journal of Allergy and Clinical Immunology: In Practice*, vol. 4, no. 1, pp. 1–8, 2016.

[10] L. Amelia, C. Riccardo, D. Chiara et al., "Omalizumab in children with severe allergic asthma: the Italian real- life experience," *Current Respiratory Medicine Reviews*, vol. 13, no. 1, pp. 36–42, 2017.

[11] K. R. Murphy and B. E. Chipps, "Tiotropium in children and adolescents with asthma," *Annals of Allergy, Asthma, & Immunology*, vol. 124, no. 3, pp. 267–276, 2020.

[12] B. Ding, Y. Lu, Y. Li, W. Zhou, and F. Qin, "Efficacy of treatment with montelukast, fluticasone propionate and budesonide liquid suspension for the prevention of recurrent asthma paroxysms in children with wheezing disorders," *Experimental and Therapeutic Medicine*, vol. 18, no. 4, pp. 3090–3094, 2019.

[13] H. Madsen, D. P. Henriksen, V. Backer, H. C. Siersted, N. Bjerring, and C. S. Ulrik, "Efficacy of bronchial thermoplasty in patients with severe asthma," *Journal of Asthma*, vol. 58, no. 7, pp. 1–10, 2019.

[14] H. Nolte and J. Maloney, "The global development and clinical efficacy of sublingual tablet immunotherapy for allergic diseases," *Allergology International*, vol. 67, no. 3, pp. 301–308, 2018.

[15] L. Brian, G. John, P. Maarja, A. Emeryk, and C. Wixon, "Fluticasone propionate 100 microg bid using a non-CFC propellant, HFA 134a, in asthmatic children," *Canadian Respiratory Journal*, vol. 10, no. 2, pp. 103–109, 2016.

[16] T. Von Schoen-Angerer, R. Madeleyn, H. Kiene, G. S. Kienle, and J. Vagedes, "Improvement of asthma and gastroesophageal reflux disease with oral pulvis stomachicus cum belladonna, a combination of matricaria recutita, atropa belladonna, bismuth, and antimonite: a pediatric case report," *Global Advances in Health and Medicine*, vol. 5, no. 1, pp. 107–111, 2016.

[17] Z. Wan, Y. Tang, Q. Song et al., "Gene polymorphisms in VEGFA and COL2A1 are associated with response to inhaled corticosteroids in children with asthma," *Pharmacogenomics*, vol. 20, no. 13, pp. 947–955, 2019.

[18] P. Gervais, I. Larouche, L. Blais, A. Fillion, and M.-F. Beauchesne, "Asthma management at discharge from the emergency department: a descriptive study," *Canadian Respiratory Journal*, vol. 12, no. 4, pp. 219–222, 2016.

[19] J. Lin, H. Tang, P. Chen et al., "Efficacy and safety evaluation of once-daily fluticasone furoate/vilanterol in Asian patients with asthma uncontrolled on a low- to mid-strength inhaled corticosteroid or low-dose inhaled corticosteroid/long-acting beta2-agonist," *Allergy and Asthma Proceedings*, vol. 37, no. 4, pp. 302–310, 2016.

[20] D. Pearlman, G. Eckerwall, J. Mclaren et al., "O021 Efficacy of budesonide/formoterol pressurized metered-dose inhaler versus budesonide alone in children (6-<12 years) with asthma," *Annals of Allergy, Asthma, & Immunology*, vol. 117, no. 5, p. S8, 2016.

[21] J. K. Gerald, B. Hallmark, D. Billheimer, F. D. Martinez, and L. B. Gerald, "Are Latino children of Mexican origin with asthma less responsive to inhaled corticosteroids than white children? - ScienceDirect," *Journal of Allergy and Clinical Immunology: In Practice*, vol. 7, no. 7, pp. 2419–2421, 2019.

[22] K. R. Chapman, R. Nils, B. Vibeke, M. Palmqvist, S. Saarelainen, and M. Briggs, "Salmeterol and fluticasone propionate (50/250 microg) administered via combination Diskus inhaler: as effective as when given via separate Diskus inhalers," *Canadian Respiratory Journal*, vol. 6, no. 1, pp. 45–51, 2016.

[23] D. Har and J. A. Bird, "Efficacy and safety of benralizumab for patients with severe asthma uncontrolled with high-dosage inhaled corticosteroids and long-acting beta 2-agonists (SIROCCO): a randomised, multicentre, placebo-controlled phase 3 trial," *The Lancet*, vol. 388, no. 10056, pp. 2115–2127, 2016.

[24] F. Arpinelli, G. Caramori, and M. S. Magnoni, "[Risk of pneumonia during long term regular treatment of stable COPD with inhaled glucocorticoids: a systematic review]," *Recenti Progressi in Medicina*, vol. 108, no. 4, pp. 175-176, 2017.

Retraction

Retracted: Advanced Cognitive Algorithm for Biomedical Data Processing: COVID-19 Pattern Recognition as a Case Study

Journal of Healthcare Engineering

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

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

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

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

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

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

References

[1] M. Elhoseny, Z. Tarek, and I. M. EL-Hasnony, "Advanced Cognitive Algorithm for Biomedical Data Processing: COVID-19 Pattern Recognition as a Case Study," *Journal of Healthcare Engineering*, vol. 2022, Article ID 1773259, 11 pages, 2022.

Research Article

Advanced Cognitive Algorithm for Biomedical Data Processing: COVID-19 Pattern Recognition as a Case Study

Mohamed Elhoseny [iD],[1,2] **Zahraa Tarek** [iD],[2] **and Ibrahim M. EL-Hasnony** [iD][2]

[1]College of Computing and Informatics, University of Sharjah, Sharjah, UAE
[2]Faculty of Computers and Information, Mansoura University, Mansoura, Egypt

Correspondence should be addressed to Ibrahim M. EL-Hasnony; ibrahimhesin2005@mans.edu.eg

Automated disease prediction has now become a key concern in medical research due to exponential population growth. The automated disease identification framework aids physicians in diagnosing disease, which delivers accurate disease prediction that provides rapid outcomes and decreases the mortality rate. The spread of Coronavirus disease 2019 (COVID-19) has a significant effect on public health and the everyday lives of individuals currently residing in more than 100 nations. Despite effective attempts to reach an appropriate trend to forecast COVID-19, the origin and mutation of the virus is a crucial obstacle in the diagnosis of the detected cases. Even so, the development of a model to forecast COVID-19 from chest X-ray (CXR) and computerized tomography (CT) images with the correct decision is critical to assist with intelligent detection. In this paper, a proposed hybrid model of the artificial neural network (ANN) with parameters optimization by the butterfly optimization algorithm has been introduced. The proposed model was compared with the pretrained AlexNet, GoogLeNet, and the SVM to identify the publicly accessible COVID-19 chest X-ray and CT images. There were six datasets for the examinations: three datasets with X-ray pictures and three with CT images. The experimental results approved the superiority of the proposed model for cognitive COVID-19 pattern recognition with average accuracy 90.48, 81.09, 86.76, and 84.97% for the proposed model, support vector machine (SVM), AlexNet, and GoogLeNet, respectively.

1. Introduction

The first infections of coronavirus disease in December 2019 (COVID-19) were recorded in a significant city in China called Wuhan. COVID-19 originates from the SARS-CoV-2 virus and is now mono of the main issues in the world. Thousands of deaths and confirmed cases worldwide are outlined from COVID-19. The high number of infections is indicative of the rapid spread between people population.

The signs of COVID-19 familiar to date involve high temperature, sore throat, cough, migraine, vomiting, muscle aches, and other symptoms. Another critical factor is the early diagnosis of COVID-19 can be mirrored in early remedy. This pandemic does not only affect the countries health, but the effects of COVID-19 are also significant (e.g., economic and psychic) [1]. So, COVID-19 is an epidemic

with a wide range of threats that should be subtended. Based on the above facts, artificial intelligence models are needed to allow the recognition of this deadly virus in the reasonable time.

On the other side, the utilization of medical photographs to detect diseases has expanded in recent years. To identify infection caused by disease, various machine vision, and image recognition techniques provide accurate and rapid results. The findings, however, must be checked by a specialist. In this manner, medical image recognition techniques can be applied as an initial diagnostic tool that hints about a potential illness.

Artificial Intelligence (AI), with countless promising reports, has been widely used in our daily operations for managing the COVID-19 epidemic. AI approaches, including deep learning, have been applied for medical

imaging to manipulate and analyze data to help clinicians and radiologists for enhancing diagnostic efficiency. Similarly, a host of research based on the automated identification of COVID-19 using deep learning algorithms have been applied [2]. AI approaches could also show their high efficiency in encouraging administrators to make smarter decisions on controlling viruses when thousands of health data are gathered by exchanging data between and across innovative countries using the recommended standards [3].

A variety of experiments have been presented to categorize COVID-19 from CT scan or X-ray images using various methodologies, such as ResNet-50, CNNs with Support Vector Machine (SVM), AlexNet, SqueezeNet, DenseNet201, VGG19, DRE-Net, and GoogLeNet. In common, it is often popular to extract various features from images and construct a collection containing details about the extracted features.

The significant contributions of this paper are summarized into three folds. In the first fold, an examination of the state-of-the-art solutions of artificial intelligence to tackle COVID-19 has been presented. As, this study introduces a systematic analysis for the most recent trials on COVID-19 using machine learning and deep learning methods; the used datasets, the tasks, and the outcomes of these trials are listed. Second, a comparative study for two pretrained models of the Convolutional Neural Network (CNN), namely, AlexNet, GoogLeNet, and Support Vector Machine (SVM) has been discussed. Also, the suggested approach is presented based on two categories of medical images: computerized tomography (CT) scanning and X-ray images. Third, a model based on the artificial neural network and the butterfly optimization algorithm for parameter adaption has been proposed.

We have used six datasets on multifaceted images of X-rays and CT scans. The utilized datasets have been used to evaluate the efficiency of the proposed model compared to the AlexNet, GoogLeNet, and the SVM models. Following comprehensive studies for both datasets of X-rays and CT scans, the proposed model has proved a highly accurate COVID-19 diagnosis. The main contributions of the current study can be summarized as follows to combat COVID-19:

(i) A review of the state-of-the-art AI solutions.

(ii) Evaluating different AI models for the COVID-19.

(iii) Designing a model for detection and prediction of COVID-19.

(iv) Using six datasets that contained three for the X-ray images and three for the CT images for practical examinations.

(v) Implementing AlexNet and GoogLeNet models as a pretrained CNN network with the SVM model for the selected datasets.

(vi) Developing a hybrid model for COVID-19 prediction.

(vii) Applying the Friedman test for comparing the proposed models and the different datasets.

The rest of the paper is organized as follows: Section 2 introduces the most recent related works. Section 3 explains the proposed COVID-19 detection model in detail, and Section 4 illustrates the experimental results and comparisons. Finally, the conclusions and future work are discussed in Section 5.

2. Literature Review

With the aid of clinical evidence and chest CT imaging, artificial intelligence (AI) methods could be applied to test the potential risks of critical cases of COVID-19. Table 1 shows the significant role of technologies that are commonly associated with AI, such as machine learning (ML), deep learning (DL), and neural networks (NN) for COVID-19 detection and diagnosis, classification, and differentiation of this epidemic from other illnesses. DL approaches, including the Convolutional Neural Network (CNN), are recommended to be the appropriate means of achieving the desired targets, especially for COVID-19 prediction and treatments. This is because that form of networks is substantially able of nonlinear modeling and has widespread use in medical image processing and diagnostics [4].

The next section discusses the proposed approach using three standard models with the hybrid ANNBOA model for COVID-19 prediction.

3. Proposed COVID-19 Prediction Model

The deep learning approaches have introduced many models in the last two years. In this section, a proposed hybrid model of the butterfly optimization algorithm with the artificial neural network (ANN) is introduced. The proposed model is compared with the pretrained AlexNet, GoogLeNet, and the SVM for diagnosing COVID-19 cases.

There are many advanced improvements for the SVM as in [22–24], but we recommended the standard method for the SVM to show the enhancements of the novel model of the deep learning approaches. Figure 1 shows the proposed framework for evaluating the ANNBOA, AlexNet, GoogLeNet, and the SVM models for COVID-19 prediction.

The proposed approach can be arranged as follows:

(1) Collecting the data using (CT, and X-ray) images.

(2) Processing data as resize images, remove noise, normalization, and others.

(3) Applying feature extraction using AlexNet, and GoogLeNet models.

(4) Classifying the output using the SVM classifier.

(5) Developing a hybrid model (BOA + NN).

(6) Comparing the hybrid model with the pretrained AlexNet, GoogLeNet, and the SVM.

(7) Evaluation and prediction the disease

The proposed approach composed of three layers. The first layer is the dataset collection from different sources of available publicity. The datasets are two categories of the medical images, the X-ray and the CT-scan images.

Datasets considered the essential factor for good training since CNN can learn how to extract significant features from

TABLE 1: AI techniques for COVID-19.

Author	Images type	AI methods	Task	Results
Nayak et al. [5]	Chest X-ray (CXR) images	AlexNet, GoogLeNet, MobileNet-V2, SqueezeNet, VGG-16, ResNet-50, ResNet-34, and Inception-V3	Classification of COVID-19 from normal cases	Accuracy of ResNet-34 is 98.33%
Shorfuzzaman and Hossain [6]	CXR images	VGG-16 network	Classification of COVID-19 cases	Accuracy is 95.6% and AUC is 0.97
Linda [7]	CXR images	A deep CNN, namely, COVID-Net	Detection of COVID-19 cases	Accuracy is 92.4%
Rahman et al. [8]	CXR images	Novel U-Net model	Automatic detection of COVID-19	Accuracy is 95.11%
Jin et al. [9]	CT images	2D deep CNN	Rapid COVID-19 detection	Accuracy is 94.98% and AUC is 97.91%
Narin et al. [10]	CXR images	Pretrained ResNet-50	Detection of coronavirus pneumonia-infected patient	Accuracy is 98%
Chowdhury et al. [11]	CXR images	AlexNet, ResNet-18, DenseNet201, and SqueezeNet	Automatic detection of COVID-19 pneumonia	Accuracy is 98.3%
Maghdid et al. [12]	CXR images and CT images	A new CNN and pretrained AlexNet with transfer learning	Effective COVID-19 detection technique	Accuracy is 98% on X-ray images and 94.1% on CT images
Gour and Jain [13]	CXR images	Multiple CNN models	Classification CT samples with COVID-19, influenza viral pneumonia, and no infection.	Accuracy is 96%, sensitivity is 98.2% and specificity is 92.2%.
Kang et al. [14]	CT images	KNN as well as NB	Automatic analysis pipeline for COVID-19	Accuracy 95%, sensitivity 93.2%, specificity 96.6%
Khanday et al. [15]	Clinical data	Multinomial naive bayes and logistic regression	Identifying pandemic with clinical text information	Accuracy is 96.2%
Sethy et al. [16]	CXR images	CNNs with the help of support vector machine (SVM)	Detecting the COVID-19 disease	Accuracy is 95.38%
Alakus and Turkoglu [17]	CXR images	CNN based LSTM, CNN-RNN	Analyzing the COVID-19	Accuracy 86.66%, precision 86.75%, recall 99.42%
Rasheed et al. [18]	CXR images	CNN and logistic regression	Diagnosis of COVID-19	Accuracy is 97.6 for CNN and 100% for LR
Gao et al. [19]	CT images	Dual-branch combination network (DCN)	Accurate diagnosis and lesion segmentation of COVID-19	Accuracy is 96.74% on the internal dataset and 92.87% on the external dataset
Goel et al. [20]	CXR images	Optimized convolutional neural network (OptCoNet)	Automatic diagnosis of COVID-19	Accuracy is 97.78%
Nour et al. [21]	CXR images	SVM	Detection of COVID-19 infection	Accuracy 98.97% sensitivity 89.39% specificity 99.75%
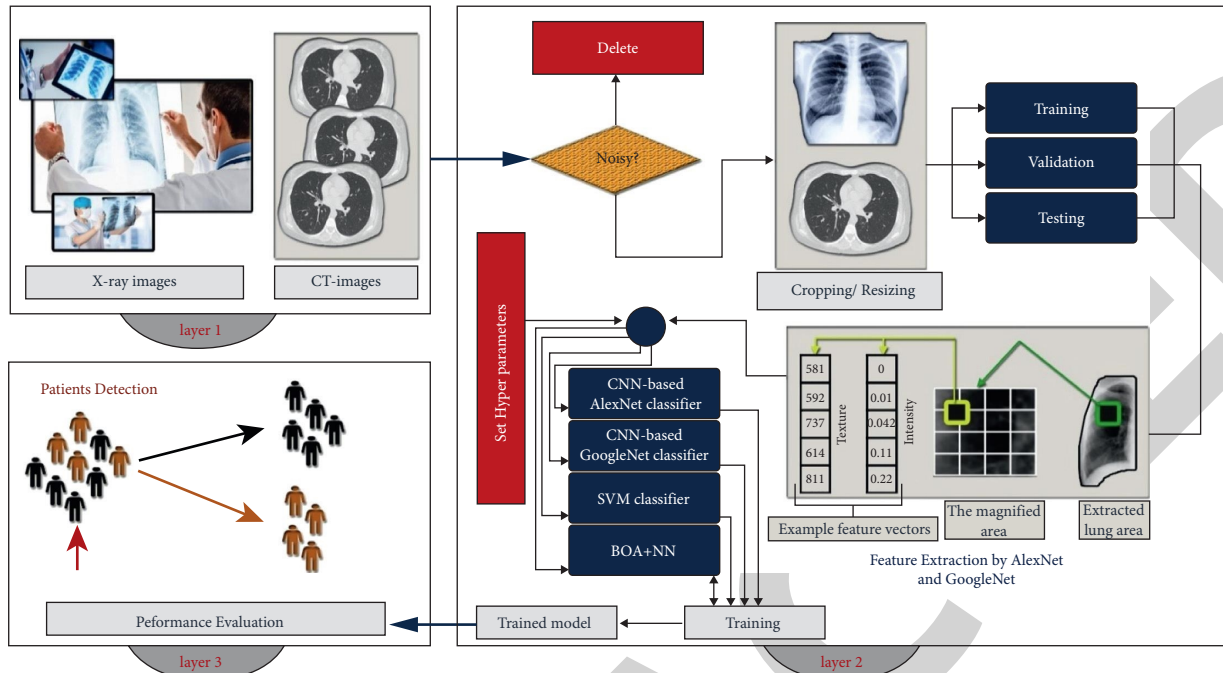
FIGURE 1: Proposed framework for evaluating different models for COVID-19 prediction.

the image and SVM classifier for detection of COVID-19 from these images. The main problem in dataset is the small number of available images about COVID-19 cases nowadays.

The quality of the developed system improves through an effective training process. The trained algorithms are confirmed by three-fold portioning of data (training-testing-validation). The training and testing process in the second layer are the most critical factor in the success of machine learning function. Validation is used to evaluate the performance and maintain the best-trained model for different hyperparameters combinations (e.g., number of iterations, the architecture, and allowable error). A final, unbiased evaluation will be performed on the test set after creating a final model based on training and validation test sets. In this paper, data are divided into training, testing, and validation by a ratio of $5:1:1$, respectively.

By comparing the activated areas of the convolutionary layers with the corresponding regions of the original images, the deep layer features of an image were examined. The activation map can have different values and was thus normalized from 0 to 1. In comparison with the original images, the strongest activation channels from COVID-19 and regular X-ray and CT images were determined. In their first convolutionary layer, convoluted neural networks detect features such as color and edges. The network can see more complicated functions in deeper convolutionary layers. Later, layers create their characteristics by combining features of older layers. COVID-19 could be challenging to distinguish from the original images of various groups of research. However, the deep layer features best explain the reason for the crash or success of a deep learning network in a demanding decision. The third layer

is the output layer with the model assessment and non-infection prediction of COVID-19 from infection. Five performance measurements, such as accuracy, sensitivity or recall, specificity, accuracy, and F1 score, assessed the performances of various networks. The proposed model could recognize the X-ray and CT images of COVID-19 from the different datasets, as they present a visual explanation of the CNN and SVM prediction and emphasizes the infected regions, and compare with the proposed hybrid model (BOA + NN) that contribute more to the classification. The proposed ANNBOA model is shown in Figure 2.

As one of the most resilient and effective machine learning approaches, neural networks (NNs) have been widely utilized to solve a variety of issues. However, selecting appropriate parameters (for example, weights) has a major impact on the accuracy of these approaches. As a result, several studies have been conducted to enhance the NN parameters. The training process of artificial neural networks, which is primarily focused on selecting the optimum combination of biases and weights, is one of the most challenging issues in machine learning. Gradient descent techniques are the most widely used training algorithms. They are, nevertheless, vulnerable to local optima and sluggish convergence in training.

The butterfly optimization algorithm (BOA) [25] is a new metaheuristic method that was recently suggested. Its idea comes from the natural food-seeking activity of butterflies. Furthermore, it has been demonstrated that BOA can tackle a wide range of optimization issues and achieving global optimal solutions. A novel classification technique based on the integration of artificial neural networks and the BOA algorithm is presented in this study.
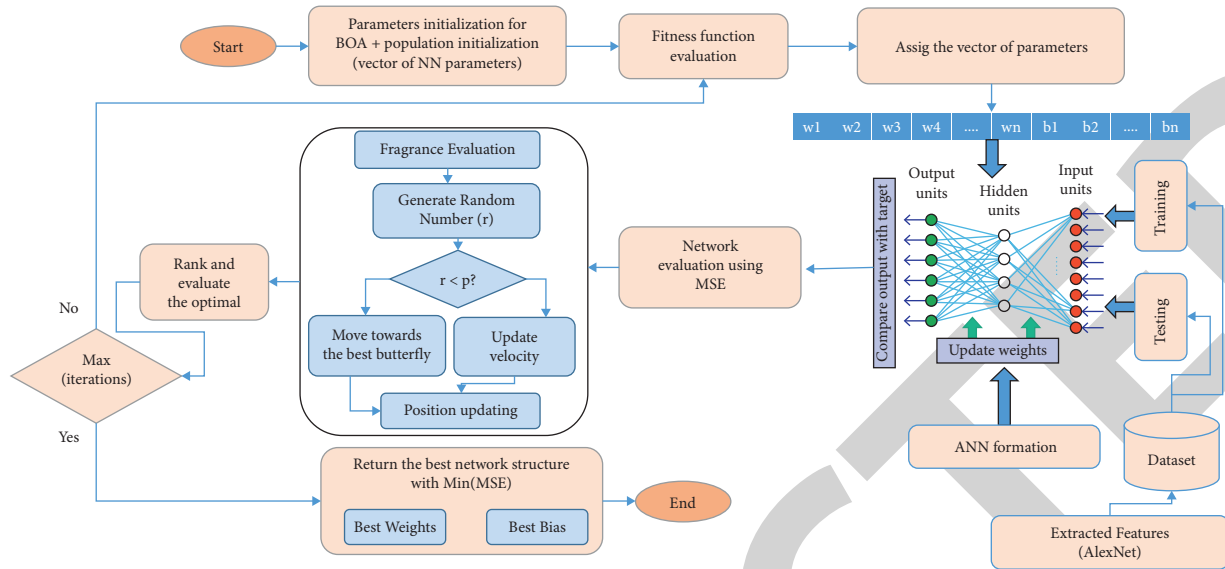
Figure 2: Proposed ANNBOA model.

The goal of the backpropagation algorithm (BP) is to improve network parameters by reducing the least square error between actual and calculated output. This section describes the BOA-BP classification technique for training the BP neural network using the BOA. The parameters for the BOA are ($c = 0.01$, a $\longrightarrow$ [0.1, 0.3], $p = 0.5$).

We execute 30 separate training and test runs to achieve significant statistical findings. We used the min-max scaling normalization approach to normalize all of the characteristics of each dataset into the [0, 1] interval, as shown in the following equation:

$$k_i' = \frac{k_i - \min_k}{\max_k - \min_{k'}}. \tag{1}$$

This procedure of normalization is essential before training since it eliminates the influence of one feature having a value in a wider range over another. The pseudocode of butterfly optimization algorithm is shown in Algorithm 1.

Although the suggested strategy has been mainly created for COVID-19 related tasks, it can apply in other medical imaging examinations. Input images are often resized for the convolution network models to maintain network architecture compatibility. The pseudocode of the utilized AlexNet and GoogLeNet models is shown in Algorithm 2.

Because the AlexNet input image is 227×227 and the GoogLeNet input image must be 224×224, we used the original image set to be redimensionable into two image sets so that both AlexNet and GoogLeNet can be used. Two matrices are necessary to train the SVM. If we have all the extracted features in a single matrix, we have an $N \times M$, which contains N as the image number and M as the feature number. The second matrix is a matrix of labels. All image labels in this $N \times 1$ matrix are being imported to tell SVM whether the given image from the set is COVID-19 or not. The last input we will use for training the SVM is these two matrices. With the SVM method, we have chosen which

features we would like to use and extracted the features with the AlexNet CNN. We can receive features of the images after we have feed images to that layer. We could then use them for training an SVM to detect COVID-19 after all the features are there. The test data are used for model assessment and prediction after training of each model.

4. Results and Discussion

In this paper, we attempt to find an SVM, transfer learning and the proposed ANNBOA so that the computers can self-detect if a particular patient is a COVID-19 using CT or X-ray pictures. Softmax was used to classify the fully connected layer with the SVM classifier in the three standard models. The entire experimentation was performed in the 64-bit Windows 10 Pro operating system, Intel(R) 16 GB RAM Core(TM) i7-8550U CPU@ 1.80GHZ 1.99 GHz processor. Each algorithm is used with MATLAB (2018a).

4.1. Dataset Description. In this paper, six datasets are used; three for X-ray and three for CT images. Table 2 gives a brief explanation of the employed datasets.

The total count of the utilized images is 18096 X-ray images, 7406 for COVID-19 cases, and 10690 for normal patients. Also, there are 10977 CT images, 6877 for COVID-19 cases, and 4100 for normal patients. The description of the dataset portioning after removing noisy images is shown in Table 3.

4.2. Results. We use a test image set to determine the threshold that can best be accurately determined after the parameters are determined. For CNN, setup is initiated by installing AlexNet and GoogLeNet pretrained networks. A sample of the network training and validation of the accuracy and the error rate (loss) for both AlexNet and GoogLeNet are shown in Figures 3, and 4 respectively.

Input ←— Total number of iteration (T_{max}), population size (**N**), objective function **f(x)**, control parameter (**a**), switch probability (**p**), sensory modality (**c**), and the power exponent (**a**).
Output ←— Optimal solution
(1) **Begin**
(2)   **For** $t = 1$: T_{max}
(3)   **For** $i = 1$: **N**
(4)   **For** $j = 1$: **d**
(5)   Update the fragrance of current search agent by Equation: $\mathbf{pf}_i = \mathbf{c} \times \mathbf{I^a}$
(6)   End for
(7)   **End for**
(8)   Find the best f
(9)   **For i** $= 1$: **N**
(10)   **For j** $= 1$: **d**
(11)   Set a random number r in [0,1]
(12)   **If** $r < p$, then
(13)   Move toward best position by Equations $\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \mathbf{F}_i^{t+1}$, $\mathbf{F}_i^{t+1} = (\mathbf{r}^2 \times \mathbf{g}^* - \mathbf{x}_i^t) \times \mathbf{pf}_i$
(14)   **Else**
(15)   Move randomly using Equations $\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \mathbf{F}_i^{t+1}$, $\mathbf{F}_i^{t+1} = (\mathbf{r}^2 \times \mathbf{x}_j^t - \mathbf{x}_k^t) \times \mathbf{pf}_i$
(16)   **End if**
(17)   **End for**
(18)   **End for**
(19)   Update the value of c and the power exponent a.
(20)   **End for**
(21) **End**

ALGORITHM 1: The butterfly optimization algorithm (BOA).

Input ←— CT and X-ray images, learning rate (U), Epoch (E)
Output ←— Trained model that classify COVID-19 images
(1) **Begin**
(2)   Preprocessing:
(3)   //$\forall x \in X$, $\exists i \in X$: i resize of x
(4)   **For** each input image
(5)   Resize images to 227×227 for AlexNet and 224×224 for GoogLeNet
(6)   Normalize images
(7)   Remove noise
(8)   **End**
(9)   DTL models $M = $ (AlexNet, GoogLeNet)
(10) Let G be a pretrained network (GoogLeNet) $\in M$
(11) Let A be a pretrained network (AlexNet) $\in$ M
(12) Let S be a set of measures: $M = $ (Accuracy, Sensitivity, Specificity, Loss, precision, F1 score)
(13) $\forall G$ & $A \in$ CNN, $\exists s | S(s) = s(CNN(x))$.
(14) **For** each M do
(15) $U = 0.001$
(16)   **For** $E = 1$ to 4 do
(17)   Update the weights
(18)   **End**
(19) End
(20) **End**

ALGORITHM 2: The AlexNet and GoogLeNet models.

To evaluate the proposed model, five measures are utilized as follows: accuracy, sensitivity, specificity, precision, and F1 score.

Figures 5 and 6 show the comparison among the different classifiers (SVM, AlexNet, GoogLeNet, and ANNBOA) for the X-ray dataset and CT-scan dataset.

Table 4 shows that all the pretrained models evaluated are successful in classifying COVID-19. SVM, AlexNet, GoogLeNet, and ANNBOA are involved in ranking images among networks trained in different X-ray and CT datasets in the problem of two classes. Figure 7 shows the mean of performance measurements for the proposed techniques.

TABLE 2: X-ray and CT-scan for COVID-19 cases and normal cases.

Type	Name	Code	COVID-19	Non COVID-19
X-ray	Extensive COVID-19 X-ray images dataset [26]	DS_1	4044	5500
	Augmented COVID-19 X-ray images dataset [27]	DS_3	912	912
	Combined COVID-19 dataset [28]	DS_5	2450	4278
CT-scan	Extensive COVID-19 CT chest images dataset	DS_2	5427	2628
	SARS-COV-2 Ct-scan dataset [29]	DS_4	1252	1229
	COVID-19 CT scans [30]	DS_6	198	243

TABLE 3: Dataset portioning.

Dataset	COVID-19				Non-COVID-19			
	Training	Validation	Testing	Sum	Training	Validation	Testing	Sum
DS_1	2880	579	575	4034	3822	785	785	5392
DS_2	3701	850	850	5401	1670	475	475	2620
DS_3	650	127	130	907	620	127	130	877
DS_4	895	177	177	1249	845	185	185	1215
DS_5	1378	407	465	2250	2436	715	927	4078
DS_6	132	30	36	198	168	39	36	243
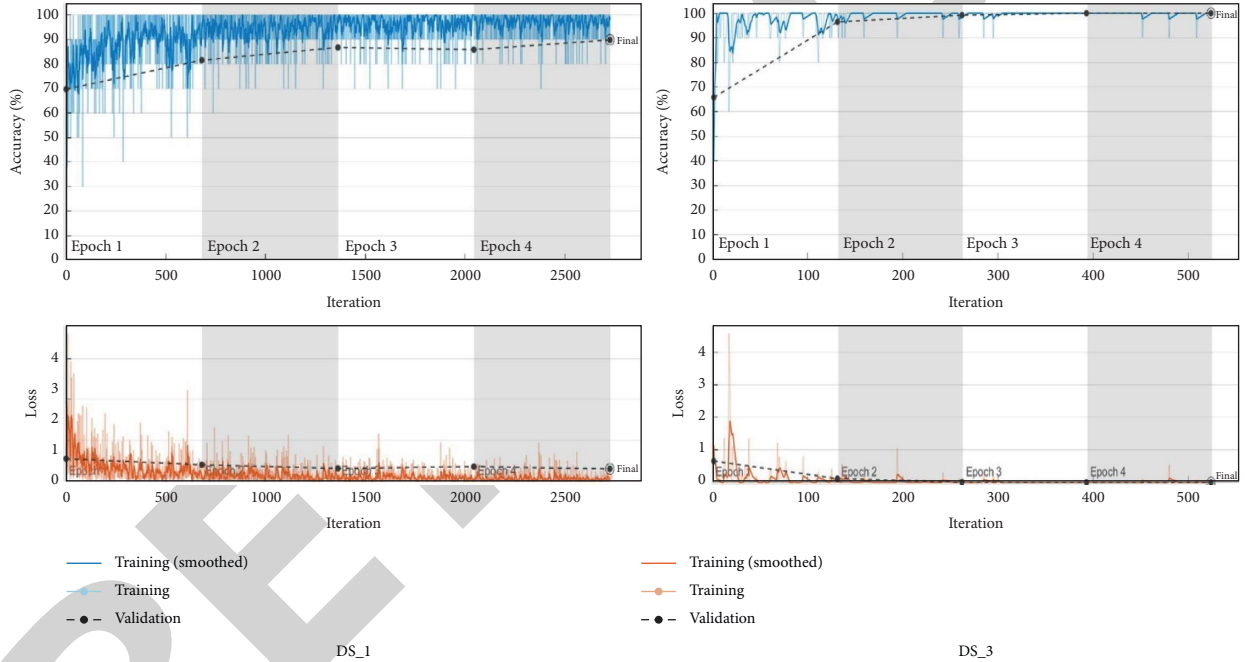
FIGURE 3: Samples from training and validation of AlexNet.

The results show the superiority of the proposed model (ANNBOA) with average accuracy 90.48%.

4.3. Results Discussion. This paper uses deep learning technologies for creating a classification network to distinguish COVID-19 from non-COVID-19 cases. As far as the network structure is concerned, AlexNet, GoogLeNet, and BOA with NN are used to extract features. The experiment showed that COVID-19 cases could be more distinguishable from others by the proposed mechanism. It is noted that from Figure 7 that the BOA with NN achieves the highest results using performance measurements (Accuracy, Sensitivity, Specificity, Precision, and F1-Score). A hypothesis test, called related-samples Friedman's two-way analysis of variance by ranks, is used to compare between proposed techniques in this paper. The null hypothesis is "the distribution of SVM, AlexNet, GoogLeNet, and BOA are the same." If asymptotic significance is less than the significance level is $\alpha = 0.05$, then the decision is rejecting the null hypothesis. As sown in Table 5, Asymp sig. <0.05, so we reject the null-hypothesis with the mean that there are statistically significant differences between 4 techniques with strong effect according to ES (Effect Size) which is used to detect impact of techniques using the following equation:

$$ES = \frac{x_{\text{Friedman}}^2}{k(N-1)}, \quad (2)$$

as $k = \#$ measurements, and $N = \#$ techniques.
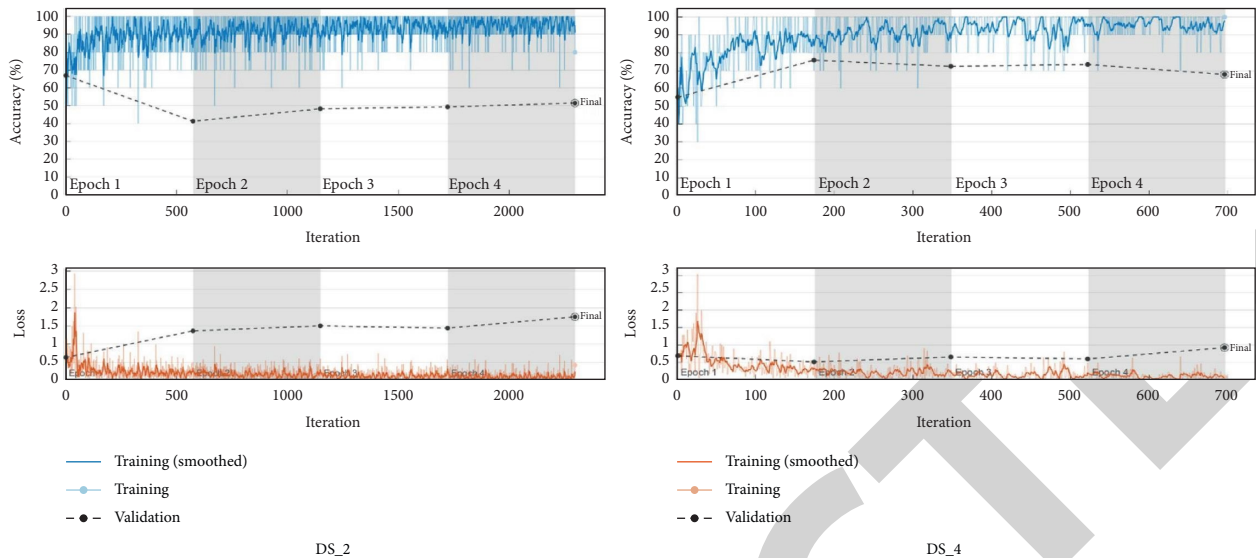
Figure 4: Samples from training and validation of GoogLeNet.
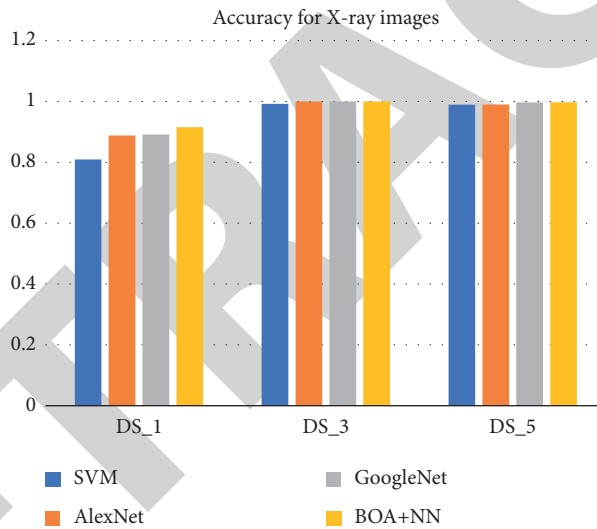
Figure 5: Comparison among the three classifiers for the X-ray images.
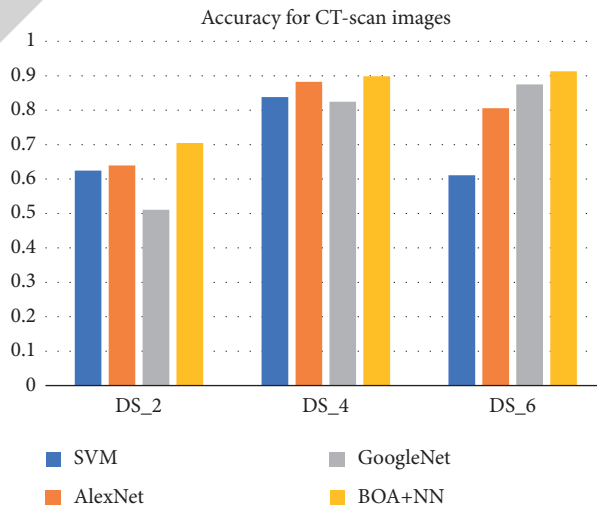
Figure 6: Comparison among the three classifiers for the CT-scan.

Table 4: Output results for SVM, AlexNet, and GoogLeNet.

Dataset	Model	Accuracy	Sensitivity	Specificity	Precision	F1-score
DS_1		80.96	79.65	81.91	76.33	77.96
DS_2		62.47	71.88	45.78	70.16	71.01
DS_3	SVM	**99.23**	99.23	99.23	99.23	99.23
DS_4		83.83	79.44	88.11	86.67	82.90
DS_5		98.94	98.95	98.94	98.21	98.57
DS_6		61.11	100	22.22	56.25	72
DS_1		88.82	92	86.50	83.31	87.44
DS_2		63.94	62.24	67.47	79.83	69.95
DS_3	AlexNet	**100**	100	100	100	100
DS_4		88.21	85.56	90.81	90.06	87.75
DS_5		99	98.50	99.29	98.79	98.64
DS_6		80.56	88.89	72.22	76.19	82.05
DS_1		89.12	90.43	88.15	84.83	87.54
DS_2		51.08	45.49	62.67	71.60	55.63
DS_3	GoogLeNet	**100**	100	100	100	100
DS_4		82.47	79.44	85.41	84.12	81.71
DS_5		99.67	99.55	99.73	99.55	99.55
DS_6		87.5	97.22	77.77	81.40	88.60
DS_1		91.54	92.35	90.93	88.58	90.42
DS_2		70.46	69.37	71.89	76.39	72.71
DS_3	BOA + NN	100.00	100.00	100.00	100.00	100.00
DS_4		89.86	89.82	89.90	88.24	89.02
DS_5		99.72	99.70	99.73	99.55	99.62
DS_6		91.30	97.22	84.85	87.50	92.11
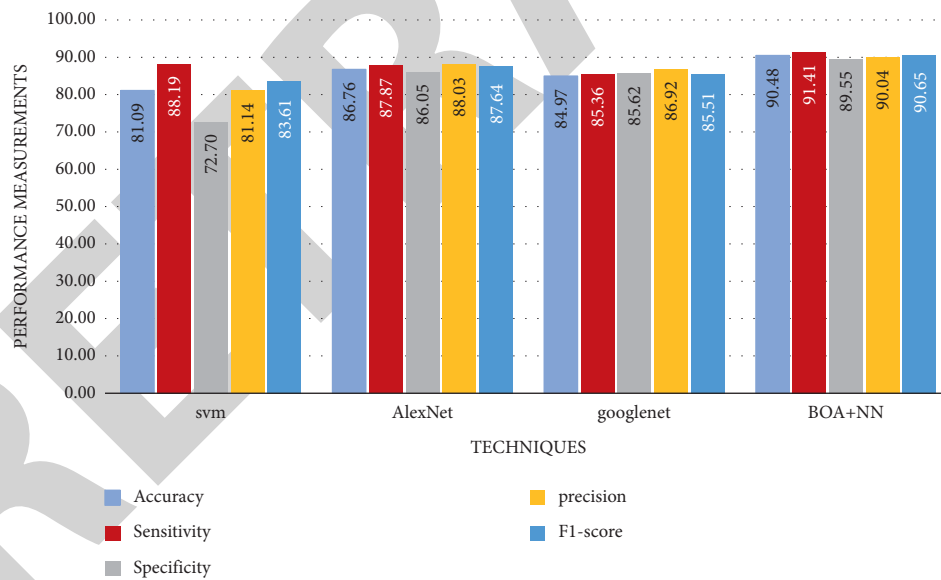
Figure 7: Mean of performance measurements for the proposed techniques.

The value of ES is compared to the standard values to check the strong of ES. In our case, ES = 0.808 that is between .70 and .90 with strong effects.

Wilcoxon Signed Ranks Test is used to list these statistically significant differences. Table 6 shows Wilcoxon Signed Ranks test between the proposed techniques. It is noted that there is significant difference between AlexNet and GoogLeNet in favor of AlexNet. Also, there are significant differences between BOA + NN and SVM, AlexNet, GoogLeNet in favor of the BOA + NN technique. So, BOA + NN is the best technique.

The correlations between the BOA + NN and the other techniques, where there are significant differences, are calculated using Spearman's correlation in Table 7.

Figure 8 shows strong positive correlation ($r = 0.9$) between BOA + NN, and SVM as P (0.037) < 0.05 in the presence of the five measurement. We conclude from the previous results that BOA + NN technique is correlated with

Table 5: Friedman test for our proposed techniques.

Model	Mean	Std deviation	Max	Mean rank	Chi-square	Asymp sig.	ES
SVM	81.35	5.63	88.19	1.40			
AlexNet	87.27	0.84	88.03	2.80	12.12	0.007	0.808 (strong effect)
GoogLeNet	85.68	0.74	86.92	1.80			
BOA + NN	90.43	0.70	91.41	4.00			

Table 6: Wilcoxon signed ranks test for SVM model.

	AlexNet-SVM	GoogLeNet-SVM	BOA–SVM	GoogLeNet-AlexNet	BOA-AlexNet	BOA-GoogLeNet
Z	-1.753^b	-1.483^b	-2.023^b	-2.023^c	-2.023^b	-2.023^b
Asymp. Sig. (2-Tailed)	0.080	0.138	0.043	0.043	0.043	0.043

b. Based on negative ranks. c. Based on positive ranks.

Table 7: Spearman's correlation.

	SVM	AlexNet	GoogLeNet	BOA + NN
SVM				
AlexNet	0.700			
GoogLeNet	−0.200	0.300		
BOA + NN	0.900*	0.400	−0.600	

*Correlation is significant at the 0.05 level (2-tailed).
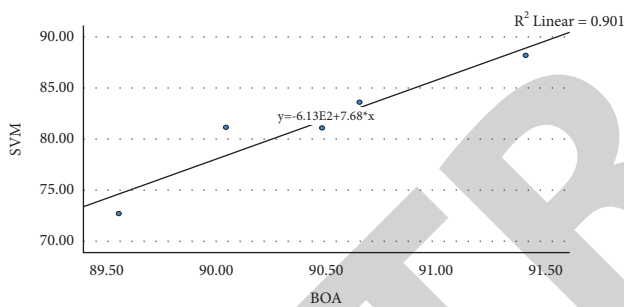
Figure 8: Correlation between BOA + NN and SVM.

SVM classifiers (positive correlation) to detect disease by high-performance measurements. There are some restrictions in this study; the first thing to do is combines the patient's contact history, travel history, initial symptoms, and laboratory examination with a clinical diagnosis of COVID-19. Second, this study limited the number of model samples. To improve accuracy in the future, the number of training and test samples in the utilized datasets must be extended.

5. Conclusion and Future Work

Coronavirus (COVID-19) recently became one of the world's worst and most acute diseases. Therefore, an automated pattern detection system should be used as the fastest possible diagnostic method to prevent the spread of COVID-19. This paper introduced a new approach for automatically screening CT and X-ray images of COVID-19 using deep learning technologies. Models can be classifying different cases for the COVID-19 and non-COVID-19 with the high accuracy among the GoogLeNet, AlexNet, and proposed BOA + NN models in some comparative data sets

so that a promising additional diagnostic tool for frontline clinical doctors may be the proposed approach. The findings revealed the dominance of BOA with NN to attain the highest accuracy of all X-ray images. In contrast, in CT images, AlexNet accomplished the highest accuracy in two datasets and GoogLeNet, in one dataset.

In future, we aim to apply different machine learning and deep learning models together with providing strategies to select the most important features using metaheuristic algorithms.

Abbreviations

COVID-19:	Coronavirus disease 2019
CXR:	Chest X-ray
CT:	Computerized tomography
CNN:	Convolutional neural network
SVM:	Support vector machine
ML:	Machine learning
DL:	Deep learning
ANN:	Artificial neural network.

Data Availability

Datasets are publicly archived at https://data.mendeley.com/datasets/8h65ywd2jr/3-https://data.mendeley.com/datasets/2fxz4px6d8/4-https://data.mendeley.com/datasets/3pxjb8knp7/3-https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset-https://github.com/UCSD-AI4H/COVID-CT.

Ethical Approval

The authors approve all ethical guidelines provided by the journal.

Research Article

Fog Computing Service in the Healthcare Monitoring System for Managing the Real-Time Notification

Ahmed Elhadad ⓘ,[1] Fulayjan Alanazi ⓘ,[2] Ahmed I. Taloba ⓘ,[1] and Amr Abozeid ⓘ[1]

[1]*Department of Computer Science, College of Science and Arts Qurayyat, Jouf University, Sakaka 72388, Saudi Arabia*
[2]*Department of Computer Science, College of Computer and Information Sciences, Jouf University, Sakaka 72388, Saudi Arabia*

Correspondence should be addressed to Ahmed Elhadad; aaelhadad@ju.edu.sa

A new computing paradigm that has been growing in computing systems is fog computing. In the healthcare industry, Internet of Things (IoT) driven fog computing is being developed to speed up the services for the general public and save billions of lives. This new computing platform, based on the fog computing paradigm, may reduce latency when transmitting and communicating signals with faraway servers, allowing medical services to be delivered more quickly in both spatial and temporal dimensions. One of the necessary qualities of computing systems that can enable the completion of healthcare operations is latency reduction. Fog computing can provide reduced latency when compared to cloud computing due to the use of only low-end computers, mobile phones, and personal devices in fog computing. In this paper, a new framework for healthcare monitoring for managing real-time notification based on fog computing has been proposed. The proposed system monitors the patient's body temperature, heart rate, and blood pressure values obtained from the sensors that are embedded into a wearable device and notifies the doctors or caregivers in real time if there occur any contradictions in the normal threshold value using the machine learning algorithms. The notification can also be set for the patients to alert them about the periodical medications or diet to be maintained by the patients. The cloud layer stores the big data into the cloud for future references for the hospitals and the researchers.

1. Introduction

Cloud computing provides a variety of IoT services, including computation resources, storage capacities, heterogeneity, and high processing that have accompanied a technological revolution. At many levels, the cloud allows for the virtualization of computational resources. Almost every aspect of human life has embraced cloud computing. Cloud computing, on the other hand, has limitations in terms of large delays, which have a negative impact on IoT jobs that demand a real-time reaction. It also does not work with industrial control systems that require a quick response time [1].

A concept of fog computing has been introduced to link IoT devices with data centers. Similar to IoT, fog computing has several applications like monitoring and analysis of data from network-connected things in real time and facilitates further actions to be taken. Fog computing is more virtualized and it can provide networking services among end devices and cloud computing data centers, but it is not entirely positioned at the edge of the network. Fog computing can be used at three levels of networking: (1) data collection from edge devices (sensors, vehicles, roadways, and ships), (2) multiple devices connecting to a network and sending all data, and (3) the collected data from the devices that should be processed in less than a second, along with decision-making [2, 3].

The IoT is considered as a dynamic global network infrastructure, in which the things with unique characteristics are unified to enable advanced services. One of the basic technologies used in IoT healthcare is the Wireless Body Area Networks (WBANs). It can acquire the signals like body temperature, electrocardiogram (ECG), electromyography (EMG), and blood pressure. These data are

transmitted to the end-users through the protocols like Wi-Fi or IEEE.802.15.4 for diagnosis and visualization. In healthcare monitoring, the remote cloud servers are in use to process and store the data from sensor nodes using cloud computing. But there are some challenges like latency sensitivity, large data transmission, and location awareness. Fog computing brings cloud capabilities closer to end-users by delivering storage, processing, and communication to edge devices, which improve mobility, privacy, security, low latency, and network bandwidth. Fog computing can ideally match latency-sensitive or real-time applications [4].

Real-Time Environmental Monitoring, Visualization, and Notification systems are considered to facilitate a new perspective in reliable data capturing, effective visualization, and handling emergency time-sensitive circumstances for Health and Safety management [5, 6]. In recent years, the population of aged people all over the world is increasing and hence complex health issues are also increasing which leads to increased clinical expenses. Remote health monitoring is vital for individuals, especially for the aged. This can help in reducing the additional expenses to be spent on hospitals. But the traditional models of healthcare monitoring are not so convenient and involve time-consuming processes. Hence, there is a demand for developing an efficient healthcare monitoring system that reduces the hospitalization of patients and also enhances the quality of life of the patients [7]. In the era of IoT, a huge amount of data is to be handled each and every second from several devices. Currently, cloud computing is used to handle such data. But the requirement of data centers which are highly expensive makes cloud computing infeasible for data processing, because of the larger distance between the data center and the sensing devices. As the healthcare data should be transmitted without any delay so as to get the proper response on time, fog computing is introduced in the field of health monitoring to improve the Quality-of-Service (QoS) [8].

Mobile health (m-health) denotes the use of mobiles or smartphones in the collection of healthcare data from the patients in real time and storing them into a cloud server using the Internet. The hospitals, insurance providers, etc. can be able to access these data from the cloud so as to give proper treatment for the patients. The availability of wearable devices and body sensor networks helps in the invention of the m-health system. Integrating m-health in a patient's environment helps in real-time analysis of the patient's health [9]. For several healthcare applications, the use of a simple sensor-to-cloud system is not feasible due to the security concern and the issue of network failure that risks the patient's health due to the delay in accessing the patient data. Using only the cloud platform for healthcare data to be processed and stored results in the delay of transmitting the data from sensors to the cloud and from the cloud to hospitals. Hence, the use of fog computing is introduced in healthcare applications in situations that require emergency responses and real-time actions [10]. The number of patients using Implantable Cardioverter Defibrillators is going on to increase. It collects the data such as device functioning and physiological parameters so as to improve the survival of the patients. Monitoring those data

in real time remotely helps in the early detection of clinical factors and promotes therapy adjustments or reprogramming the devices before the patients are hospitalized. Remote monitoring of health is highly effective compared to the traditional one-on-one consultation since it is highly reliable, safe, cost-efficient, and quick responsive regarding any failures, minimizes the hospital appointments, and reduces the unwanted shocks regarding their health condition [11].

The applications of IoT paves a way for the implementation of ambient-assisted living systems for providing assistance in health monitoring and even in the day-to-day activities of an individual. The patients monitored in hospitals, especially in Intensive Care Units (ICUs) by medical assistants sometimes, causes errors since the errors are inevitable for humans. Hence, it is essential to develop an IoT-enabled health monitoring system for patients needing permanent support, in any case, to accurately collect the health parameters to provide the right clinical support on time and to prevent unnecessary costs and efforts on improper monitoring [12]. Fog computing is formed by extending cloud computing to the edge of the network. Fog computing is not a substitute for cloud computing, but it will act as a complement to it. The heavy network traffic nowadays causes a huge burden in bandwidth while transmitting a large amount of data to the cloud, unbearable latency, and degraded services to the end-users. Hence, it is important to redesign the computing patterns so as to meet the necessities of the big data era [13]. A detailed systematic literature survey has been given in [14] which includes the motivations, limitations, and future recommendations in this field of IoT healthcare systems using fog computing. This survey describes that fog computing can be well suited for real-time applications that require high response time and low latency. It is clear from this review that fog computing decreases the latency compared to cloud computing, which is very much essential in the field of healthcare in real time.

The objectives of the proposed fog computing paradigm are as follows: it may reduce latency when transmitting and communicating signals with faraway servers, allowing medical services to be delivered more quickly in both spatial and temporal dimensions. As a result, machine learning is used to process the data, which then becomes the outcome of interpretation. The proposed tools and technologies are explained in detail, including how they can work in tough circumstances and how they can be implemented in areas where IoT-driven healthcare services are most required. The rest of the paper is organized as follows: Section 2 presents related works undergone during the research process, Section 3 presents the proposed methodology, Section 4 gives results and discussion followed by Section 5 with a conclusion, and finally, the reference paper used in this research has been listed out under reference section.

2. Related Work

An IoT-based healthcare system that uses the fog computing concept is given in [4]. The efficiency of using fog computing in the field of health monitoring has been demonstrated

based on the use of bandwidth, QoS, and notification alerts. Also, a case study on using ECG features extracted from the clinical data at the edge of the network in the implementation process has been given at last. This paper collects the data such as temperature, the humidity of surroundings, location of the patients, SpO2, Electro Encephalography (EEG), ECG, and EMG, and these data are transmitted using the communication protocols, namely, Wi-Fi, Bluetooth, ZigBee, or 6LoWPAN. The advancement in technologies allows significant use of wearable devices in the field of healthcare to be deployed with IoT. Instead of using external sensors for monitoring, the use of devices that are employed in an individual's day-to-day activities is encouraged, for example, mobile phones. Since mobile phones are devices that have various sensors built-in, like gyro meter, heartbeat sensor, accelerometer, etc., remote monitoring of patients can be supported only by using the mobile phones themselves. But the use of fog computing is necessary so as to facilitate computing, networking, and cloud data storing [15]. The security breaches include impersonation, data breaches, data integrity, eavesdropping, and collusion. The clinical data obtained from the sensors or IoT devices are highly susceptible to security threats. These data would be highly confidential and should be accessed only by authorized individuals. The security breaches make those data to be accessed by unauthorized individuals revealing its confidentiality. Hence, a solution for solving the security breaches in healthcare IoT systems has been given in [16], which provides a fog-based decentralized healthcare architecture with multiple virtual machines having the fog nodes distributed geographically for data management and processing. Each virtual machine is responsible for a specific type of data and request.

A framework on Resource Preservation Net for systems in the emergency department using Petri net has been developed combined with custom cloud computing and edge computing technologies. The Resource Preservation Net is well suited for real-time applications in which the patient's length of stay (LoS), average waiting time, and resource utilization are considered vital performance indicators. Hence, in this framework, it is indicated that the LoS, average waiting time, and resource utilization are considerably improved [17]. The IoT concept which is the combination of sensors, smart devices, and the Internet and the middleware concept which is the combination of sensors and IoT devices have been described in [18]. This paper suggests that there will be a crucial motivation for the integration of middleware in e-health systems. The integration of middleware platforms in e-health systems is to improve the process of monitoring the general health of a patient or an individual. On-site research on IoT-based mobile medical information systems has been described in [19]. This research includes a questionnaire survey, evaluation of people's opinions which includes the disease diagnosis by hospitals, the treatments and the cost of treatment procedures, and hospital's information systems. With the help of the results obtained from the survey, the requirement analysis and design of the mobile medical information system had been accomplished.

E-health and m-health provide hospitals, students, healthcare professionals, and researchers with various services like disease diagnosis, risk prediction, treatment analysis, health monitoring, education, and machine learning model development. In [20], a literature survey of the effective use of IoT in the healthcare domain has been carried out, followed by the development of a semantic model of e-health and named "k-Healthcare" for accessing the healthcare data by the patients effectively using smartphones. Since most of the existing e-health and m-health models do not use smartphone sensors for sensing and transmitting health data, the k-healthcare uses smartphones for sensing and transmitting the data shown in Figure 1. It consists of four layers of architecture, namely, the sensor layer, network layer, Internet layer, and finally a service layer. Each layer provides unique functions to facilitate efficient services to adjacent layers. A case study on the challenges faced in IoT because the models being limited to data centers cannot satisfy the needs of several IoT applications is given in [21]. This study focuses on 3 requirements such as mobility, actuation, and control of reliability and scalability, particularly where a large span of areas is covered and for real-time decision-making scenarios. The method of applying IoT, cloud computing, and fog computing in home-based hospitalization has been given in [22]. In addition to this, this paper also describes the concept of edge computing in remote health monitoring of patients and telehealth which supports the communication between healthcare professionals or doctors and the patients. This study indicates that the integration of fog computing and cloud computing in IoT healthcare technology provides quality services to patients. This method also proves that it offers high safety, reliability, and cost-effectiveness.

An IoT-based e-health system using fog computing offers patient-centric healthcare instead of clinic-centric treatment in which all authorities like patients, hospitals, and services are interconnected with each other seamlessly. The promises and the challenges faced during the transition of clinic-centric to patient-centric approaches have been demonstrated and described in [23]. This study described the benefits of applying IoT technology in e-health systems, which are as follows: all-encompassing, which means that it can be applied for different types of purposes like healthcare, exercising, beauty, safety, etc.; integrating with distinct technologies irrespective of its complexity seamlessly; analysis and processing of big data; personalizing the services; for instance, it can foresee and predict the health issues that would be caused in future so as to make the patients take necessary steps for prevention or curing; lifetime monitoring; easy to use; reduced cost; accessible at anywhere and anytime; more number of doctors could be involved if needed; online assistance; and better efficiency and international collaboration. A fog-centric smart home-based patient health monitoring system using IoT is demonstrated in [24]. The major objective of this system is to monitor the patients who need intensive care in their home itself remotely. This model consists of 5 layers, namely, data acquisition, event classification, information mining, decision-making, and cloud storage layers. Each layer is
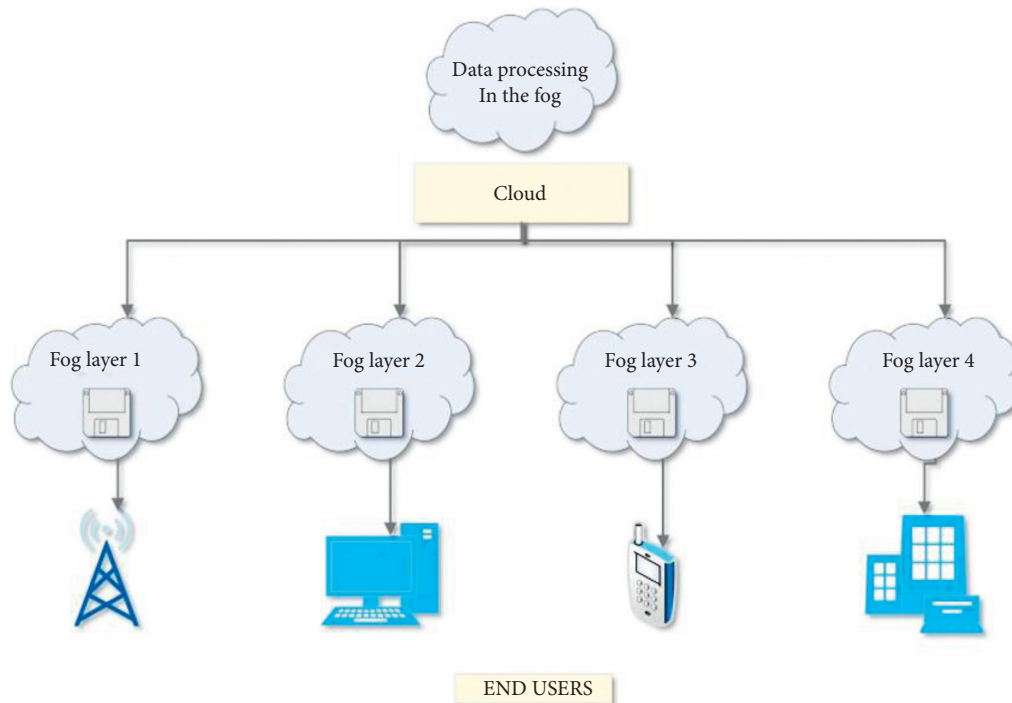
FIGURE 1: Fog distributions to end-users by cloud.

responsible for unique functions so as to provide better services to the adjacent layers. This paper offers services like home-based patient monitoring, event classification using fog computing to provide real-time response, temporal mining of health-related data from patients based on event triggering in the cloud layer, and real-time notification-based decision-making and data transmission to caregivers and hospitals in patient's unsafe conditions. A fog assisted Wearable IoT (wIoT) system for end-to-end analytics has been implemented in [25]. The fog computing can help getting efficient wIoT system. The wIoT devices are connected in one end and the cloud at the other end. In this paper, a prototype made of a middle layer, which is the smart for gateway using Intel Edison and Raspberry pi, has been developed. This prototype is designed to support data conditioning, filtering, analysis, and transmitting appropriate data to cloud for storing it for a long time and temporal variability monitoring. This prototype was tested with smart e-textile gloves system and observed that the conversion of real-world data to useful analytics with the help of knowledge-based models increases the usability of this system. Hence, it can provide better end-to-end interaction of wearables and cloud [26–29].

3. Preliminaries

3.1. Fog Computing. Fog computing lies between the devices and cloud computing. This is made possible by introducing a new processing unit between the cloud and the user to enhance reliability, energy efficiency, and privacy maintenance and reduce latency. In addition to the advantages of cloud computing, fog computing also facilitates computing power, networking capability, storage capacity, and real-time data analysis. Fog computing also facilitates real-time notification systems on users' devices. Unlike cloud computing, fog computing reaches the users in no time, but latency is the main issue in cloud computing. But the fog computing is not as much powerful as cloud computing, since it makes use of low-end computers, mobile phones, and personal devices only.

The data from all devices have been transmitted to the fog computing devices which then go via various layers such as (1) Data Center/Cloud layer, (2) Core Domain/Network layer, (3) as Edge Domain layer, (4) Smart Sensors layer, and finally, (5) Smart Monitoring layer as given in Figure 2. To overcome data corruption, a security layer exists between the end-user device and the computing devices. The data are then transmitted to the devices of fog computing for processing. This step is called preprocessing. These data are not big data. Only a small amount of data will get processed by fog computing devices. An additional advantage of fog computing is that there is a temporary storage layer that can store the processed data, which can be sent later to cloud computing for processing further.

The maintenance of a fog computing network will be highly complicated. The maintenance persons will take care of the network if there occurs any fault in the network during data transmission. The major advantage is the fact that the computational service could learn and adapt the possible outcomes obtained from fog computing.

3.2. Fog Computing in Health Monitoring. Fog computing is a novel standard in the system of healthcare monitoring. Fog computing facilitates data processing without reducing its quantity and also prevents network congestion. The remote

FIGURE 2: Various layers of fog computing.

healthcare system makes use of body sensors that are embedded into the patient body or perfectly positioned on the body to monitor the health status of the patients by gauging certain signs and to assist them in getting exact solutions. A specific device is used in collecting the data from those body sensors and also supports the communication between the sensors and the monitoring devices.
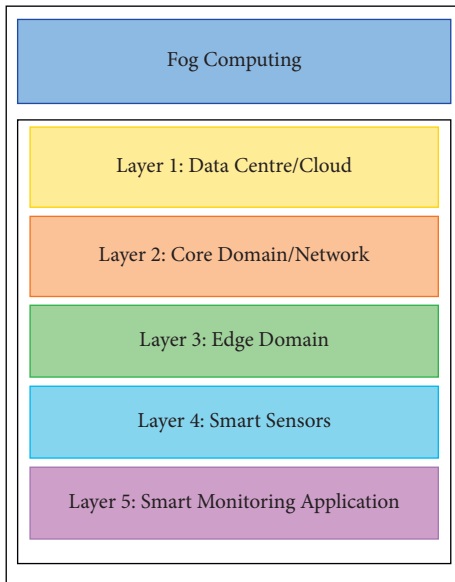
4. Proposed Methodology

Healthcare issues are becoming more prevalent in overpopulated countries, such as India, as the population grows and the need for medical assistance grows. The population's need for high-quality care is increasing, while treatment costs are decreasing. Technology has progressed to the point that health can be monitored remotely via a machine, which is more dependable than manual monitoring. It can assist in reducing the time spent on individual personal training and increasing the reliability of advanced machinery. Wearable devices, which track everything and everyone in every imaginable way, are one form of fog computing that is disrupting our day-to-day lives on a daily basis. The devices, which are wearable by people, contain a variety of sensors. These devices maintain track of a variety of human activities. Wearable devices provide a new edge for healthcare, but the potential of the Internet of Things is growing in every discipline, not just medical sciences. These wearable devices make life easier and more comfortable for people by tracking health data, as well as safety and security and their associated technologies.

There is an existing system for triparty, one-round key authenticated agreement, employing fog computing facilities, and it used bilinear pairing encryption to generate a session key between the persons to assure safe communication. A new computational framework for remote real-

time monitoring, sensing, and scaling high-performance computing for prognosis and diagnosis has been proposed in this paper.

The proposed healthcare monitoring framework consists of three layers, namely, (1) sensor network layer, (2) fog layer, and (3) cloud layer. The architecture of the proposed fog-based health monitoring framework is given in Figure 3. These layers are described as follows.

4.1. Sensor Network Layer. In order to sense the temperature, pulse rate, and blood pressure of a patient, sensors such as temperature sensor, ECG sensor, and blood pressure sensor are used along with a wearable device in which these sensors have been embedded so as to be comfortably worn by the patients. All these sensing devices form the sensor network layer. This layer senses the patient's health conditions which include the patient's body temperature, heart rate, and systolic and diastolic blood pressure and collects the data using the sensors. These data are then transmitted to the fog computing devices through wired or wireless communication protocols.

4.2. Fog Layer. The fog layer comprises several distributed nodes. These nodes are called gateways. The gateway is a device that facilitates storing, computing, and network connection that is distributed near the sensors. These sensors are responsible for recording the events and producing the data. This layer facilitates four operations: data reception from sensors, analysis of those data for health-related decision-making, notifying the caregivers, and storing all these data in the cloud. In order to improve the reliability of the system, reduce latency, prevent network connectivity issues, and increase the speed of decision-making processes, a local data processing system has been implemented in fog computing to improve its intelligence. The properties of the fog layer are given as follows:

(i) Data collection: the data from various sensors like temperature sensors, ECG sensors, and blood pressure sensors, which are connected to the wearable device, are collected for further analysis and decision-making process. After collecting the data, filtering, noise removal, and preprocessing are performed.

(ii) Data security: the data of patients obtained from the sensors are to be transmitted from gateway devices to the cloud. Hence, there is a need for the security and integrity of such data to be considered in the design of a healthcare monitoring framework. Thus, a process of encryption and watermarking has been implemented so as to secure the proposed framework. In the encryption process, the data are converted to an unrecognized format so as to protect them from unauthorized persons. In the watermarking process, the data are hidden behind an image without any distortion in its credibility or visibility [30–34].
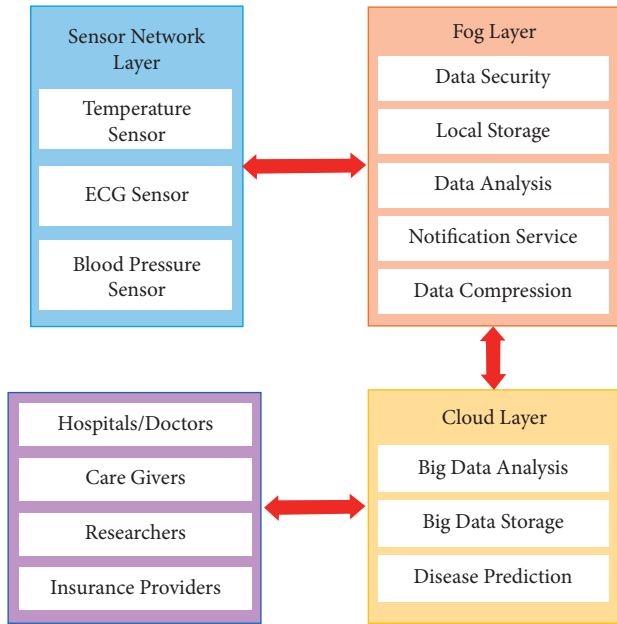
Figure 3: Architecture of the fog-based health monitoring framework.

(iii) In the proposed system, the patient's data are hidden under the patient's face image by watermarking while storing in the fog server or transmitting to the cloud server. After that, the watermarked image has been converted into a cipher image using an appropriate encryption algorithm. This encrypted data cannot be read by unauthorized persons. The data needs to be stored under the patient's image; this is done because the data cannot be opened unless permission is given by the authorized person. The data must be hidden because these data are used to be stolen and people improve their intelligence in blackmailing by these collected data. Figure 4 shows the process of watermarking and encryption of data into a face image. Meanwhile, the author states that the storage space is a concern and thus it is necessary to compress the data. The transmission of the patient's image for each collected data turns out to be an excessive overheard. By adjusting the parameters in the model and changing the evolution mechanism of the network, the robustness of fog computing can be improved.

With this growth, fog computing scale with its related edge computing paradigms, such as multi-access edge computing and cloudlet, is seen as promising solutions for handling the large volume of security-critical and time-sensitive data that is being produced by the IoT.

(i) Local storage: there will be a local storage in the gateway which serves as a local repository for storing the data for analysis and security processing. In addition to this, the data will be stored temporarily in the local storage before transmission, and

even if there occurs a network cut-off, the data will be temporarily stored there till the network is resumed.

(ii) Notification service: the notification system sends notifications in tough situations of patients to the doctors or caregivers. Notification can also be sent to the patients to take proper medication on time or alerting their food time and what to eat and some other instructions to be followed by the patients.

(iii) Data analysis: the data collected from the temperature, ECG, and blood pressure sensors are stored in the fog layer, where they are analyzed for any emergency situations based on age, height, weight, and disease using a machine learning approach. If this analysis shows any contradictory situation, then a notification will be sent to the authorized persons and transmit the data through the cloud. If the analyzed data is normal, then no action will be taken further, and no data will get transmitted.

(iv) Data compression: the patient's data obtained from several sensors are stored in the cloud storage. These data could be used for disease prediction, risk analysis, and long-term use for researchers. Hence, a large amount of data would be stored in the cloud which makes the network congested, increases latency, and needs more storage space. Data compression is the best way to solve such a problem. There are two types of compression available: lossless compression and lossy compression. The proposed system uses lossless compression since the loss in medical images gives distortions in the data and leads to improper diagnosis of diseases and risk predictions. The lossless compression involves two steps which include changing the format of the image and then removing coding redundancy using an entropy encoder.

4.3. Cloud Layer. The cloud layer is composed of the distributed resources, the repositories, and the servers. The cloud manager is responsible for managing all the devices connected to the cloud layer and facilitates the patient's data reception, processing, and storing. These data can be used for analyzing the patient's health history and current status. The properties of the cloud layer are given as follows:

(i) Data storage: after the process of data analysis of the fog layer, the analyzed data are transmitted to the cloud layer, which provides a large storage space for storing the patient healthcare data for future analysis for caregivers, doctors, hospitals, and insurance providers.

(ii) Data analysis: the patient's health data, which includes the images of diseased parts, details of symptoms, treatments, and therapy plans, stored in the cloud, are subjected to analysis for future research purposes in the field of clinical decision-making. Several machine learning approaches and data visualization techniques could be
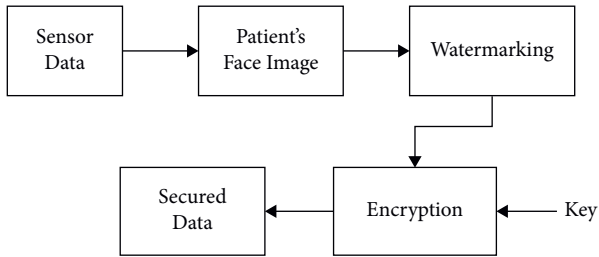
Figure 4: Watermarking and encryption process.

implemented so as to obtain more understanding from these data.

(iii) Disease prediction: the possibility of which disease the patient will get affected in the future can be predicted by the age, height, weight, and patient's hereditary. With respect to the correlation of the vital signs, the percentage of disease expectation can be predicted by using machine learning Algorithm 1.

5. Discussion

The data flow of the proposed health monitoring framework is given in Figure 5.

(i) The data from sensors are collected and are subjected to analysis.

(ii) The unwanted noise, electromagnetic interferences, and inappropriate attachments of sensors are filtered. The filtered data are analyzed so as to diagnose the current health situation of the patient.

(iii) After data analysis, if any contradictions are found in the sensed health parameter values, then a notification will be sent to the doctors or caregivers of the patients who are in need of assistance regarding their health status.

(iv) The data collected and processed are allowed to be stored in the cloud for accessing them in the future for researchers, healthcare professionals, or the patients if needed. The stored data will be watermarked under the patient's image and are encrypted to preserve the security of the patient's data during transmission and storing.

(v) These data are compressed so as to minimize the storage space required since, in the big data era, storage space is a high concern.

(vi) The data are temporarily stored locally in the fog, in case of any network failure occurs.

(vii) The most common way of getting data out of smart sensors is to use a bridging device known as a gateway in each room. A gateway receives data from the sensors and makes it useable. Data is transmitted from the sensors to the gateway wirelessly.

(viii) A smart electrical grid is a use case for fog computing. In order to function efficiently, smart cities must respond to rising and falling demands, reducing production as needed to stay cost-effective. This means that smart grids demand real-time electrical consumption and production data.

(ix) Fog computing works by utilizing local devices termed fog nodes and edge devices. Raw data is captured by IoT beacons. This data is sent to a fog node close to the data source. This data is analyzed locally, filtered, and then sent to the cloud for long-term storage if necessary.

For IoT designers, fog computing serves as the best choice due to its following features:

(i) fog computing can provide better delay performance since the position of fog resources will be between the smart devices and the cloud data centers.

(ii) fog computing requires micro centers with limited processing, communication, and storage facilities; the cost of distribution of micro fog centers near the end-users will be very less when compared to the cloud data centers.

(iii) The IoT systems are highly scalable in fog computing, which means that if the end-user increases, the micro fog centers could also be deployed more to compensate for the increase in load. The increase in the deployment of the cloud data center is not much possible, since the cost will be very high.

(iv) Fog computing provides replicated and robust services.

(v) The fog resources can serve as mobile cloud since the fog resources are deployed closely with the end-users.

(vii) The real-time services can achieve high performance with fog computing.

(viii) The fog resources can be interoperable with several cloud service providers. Hence, fog computing and its resources are highly standardized.

(ix) Data aggregation can be performed by fog resources for sending the data that are processed partially as an opposition with the raw data to cloud data centers for processing it further.

Hence, it is evidenced that fog computing has the ability to enhance the IoT device performances since a part of high-level services provided by cloud computing can be performed by local resources.

6. Evaluation

Test results on the customized ECG equipment were performed during the sample assessment processes with various functioning conditions, including sampling rates and channel count. The number of stations must be between 6 and 12, as well as the sample rate must be between 550 Hz

---

(i) Input: n number of health attribute data with threshold predetermined
Output: patient's current health state
(i) Attribute determination
(ii) Compute the abnormality
(iii) IF value > threshold
(iv) Patient is unsafe
(v) Notify the hospitals/doctors/caregivers
(vi) Transmit appropriate data to cloud for further analysis
(vii) ELSE
(viii) Patient is safe
(ix) RETURN Patient's health status

ALGORITHM 1: Predicting patient health status and notifying the authorities.

FIGURE 5: Flow diagram for the proposed system.

and 1 kHz. The data for the longest recorded duration utilizing the software's embedded 4 GB NAND Flash Storage are summarized in Table 1. In both operating conditions, as indicated in Table 2, the duration of the devices in terms of capacity charge is also examined, in particularly the real resource usage of the most precise sampling frequency of 1 kHz for 12 channels.

The Android Applications were assessed by determining the overall storage space required to store an ECG tracing with several channels and two different sampling speeds (1 kHz and 550 Hz) in EDF + formats. The entire period of all ECG traces was one minute. It is feasible to reduce the storage amount required by using compression technologies, but this would have an instant impact on battery capacity. In two different settings, the data communication rate of the ECG equipment was also tested as illustrated in Table 3.

The wearable ECG equipment could send information to a mobile phone in two modes: downloads mode, which sends information from the ECG software's memory, and monitoring mode, which sends data from the present recording in real time. The wearable ECG equipment could be in either an inactive or active recording phase when there is data transmission from an ECG. In the latter instance, the technology of the wearable ECG equipment restricts the information transfer rates to the mobile phone in order to maintain the efficiency of the active recorded collecting data.

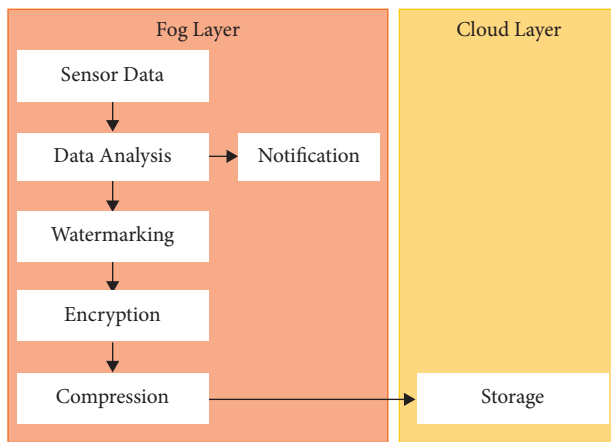The information transfer rates also differ depending on whether the ECG equipment saves the information when transferring it or if the information is delivered straight without being stored in the internal storage. The transmission of data occurs in the first scenario after the information is recorded in the storage (each admission in the storage occurs only after a full information site has been loaded), and the data is read from their saved location to establish transmission packages with a predetermined number of samples described by the device software. In the second scenario, data transfer occurs immediately after the ADC inverters have sampled the data and acquired a predetermined number of samples. The sampling frequency, the number of channels employed in the recorder, and the preset maximum number of transfers each second are used to determine the predetermined packets for information transfer.

Tables 4 and 5 illustrate the information read rates from the ECG equipment to the mobile applications for various ECG device setups, as measured by the mobile phones. We should point out that we not only transfer and measure information packets during transmitting data but also regulate and acknowledge packets, energy capacity phase report packets, and reconnection packets (in the event of a separation). These packets are frequently transmitted, but they can have an impact on the theoretically possible throughput in the communications among the ECG and the cellphone.

Any authorized users with accessibility to the Spark-Heart prototype's online site, which is accessible via the clouds, can send instructions to the Smartphone Applications and Wearable Gadgets, which pass through the system's specific interfaces. We estimated the maximum and average response time of simple condition demands (for example, ECG trace counts, remaining memory, and battery level) toward the ECG machine, as well as the maximum and average time the structure takes to make a problem for a full document of the existing ECG device state, to evaluate the end-to-end response time of such a request. Table 6 presents a summary of the results.

Table 1: Maximum length of recording.

Sample range (kHz)	Channels	Duration (hrs)
1	6	59.73
550	6	120.53
1	12	30.91
550	12	15.10

Table 2: Maximum operation lifetime.

Operation	Channels	Sampling rate (kHz)	Device lifetime (hrs)
Real-time data transfer	12	1	7.51
Non-real-time data transfer	12	1	19.62

Table 3: ECG trace parameters.

Channels	Sampling rate (kHz)	Total size (KB)
6	550	376.525
6	1	735.4
12	550	1.506
12	1	735.4

Table 4: Direct transmission of an ECG trace.

Channels	Sample rate (kHz)	Average throughput (kbps)	Storing memory
6	550	11.620	Yes
6	550	15.379	No
6	1	22.929	Yes
6	1	27.083	No
12	550	36.844	Yes
12	550	47.544	No
12	1	76.099	Yes
12	1	97.699	No

Table 5: ECG transmission storage.

Channels	Sample rate (kHz)	Average throughput	Sampling
6	550	13.158 kbps	Yes
6	550	1.566 Mbps	No
12	1	78.665 kbps	Yes
12	1	640.198 kbps	No

Table 6: Response time maximums and averages.

Command type	Maximum response (ms)	Average response (ms)
Simple report	49.829847	22.499428
Completed report	139.450269	107.994685

7. Conclusion

This paper proposed a fog-based health monitoring framework, which makes use of fog gateways for clinical decision-making based on the data collected from the sensors like temperature sensor, ECG sensor, and blood pressure sensor, which are embedded into a wearable device so as to measure the temperature, heart rate, and systolic and diastolic pressure of a patient. These data are then secured by watermarking and encryption processes and are temporarily stored in the fog server until receiving proper network connectivity and transmitting them to the doctors or caregivers via the cloud in case of any emergency situations. This increases the utility of the proposed system. The huge amount of data collected from these sensors is compressed and stored in the cloud server for future references of hospitals and researchers. Finally, in this paper, the benefits of choosing fog computing services in healthcare monitoring and clinical decision-making have been discussed.

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 to report regarding the present study.

Acknowledgments

References

[1] N. Mani, A. Singh, and S. L. Nimmagadda, "An IoT guided healthcare monitoring system for managing real-time notifications by fog computing services," *Procedia Computer Science*, vol. 167, no. 2, pp. 850–859, 2020.

[2] C. S. Nandyala and H.-K. Kim, "From cloud to fog and IoT-based real-time U-healthcare monitoring for smart homes and hospitals," *International Journal of Smart Home*, vol. 10, no. 2, pp. 187–196, 2016.

[3] A. I. Taloba, R. Alanazi, O. R. Shahin, A. Elhadad, A. Abozeid, and R. M Abd El-Aziz, "Machine algorithm for heartbeat monitoring and arrhythmia detection based on ECG systems," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 7677568, 9 pages, 2021.

[4] T. N. Gia, M. Jiang, A.-M. Rahmani, T. Westerlud, and P. Liljeberg, "Fog computing in healthcare internet of things: A case study on ecg feature extraction," in *Proceedings of the IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing*, pp. 356–363, Liverpool, UK, October 2015.

[5] A. Kiani, A. Salman, and Z. Riaz, "Real-time environmental monitoring, visualization, and notification system for construction H&S management," *Journal of Information Technology in Construction*, vol. 19, no. 1, pp. 72–91, 2014.

[6] F. Alanazi, A. Elhadad, S. Hamad, and A. Ghareeb, "Sensors data collection framework using mobile identification with

secure data sharing model," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 5, 4258 pages, 2019.

[7] H. H. Nguyen, F. Mirza, M. A. Naeem, and M. Nguyen, "A review on IoT healthcare monitoring applications and a vision for transforming sensor data into real-time clinical feedback," in *Proceedings of the IEEE 21st International conference on computer supported cooperative work in design (CSCWD)*, pp. 257–262, Wellington, New Zealand, July 2017.

[8] M. Al-Khafajiy, L. Webster, T. Baker, and A. Waraich, "Towards fog driven IoT healthcare: challenges and framework of fog computing in healthcare," in *Proceedings of the 2nd international conference on future networks and distributed systems*, pp. 1–7, Amman, Jordan, June 2018.

[9] S. H. Almotiri, M. A. Khan, and M. A. Alghamdi, "Mobile health (m-health) system in the context of IoT," in *Proceedings of the IEEE 4th international conference on future internet of things and cloud workshops (FiCloudW)*, pp. 39–42, Vienna, Austria, August 2016.

[10] A. Paul, H. Pinjari, W.-H. Hong, H. C. Seo, and S. Rho, "Fog computing-based IoT for health monitoring system," *Journal of Sensors*, vol. 2018, no. 1, Article ID 1386470, 7 pages, 2018.

[11] M. Bertini, L. Marcantoni, T. Toselli, and R. Ferrari, "Remote monitoring of implantable devices: should we continue to ignore it?" *International Journal of Cardiology*, vol. 202, no. 1, pp. 368–377, 2016.

[12] I. Chiuchisan, H.-N. Costin, and O. Geman, "Adopting the internet of things technologies in health care systems," in *Proceedings of the International Conference and Exposition on Electrical and Power Engineering (EPE)*, pp. 532–535, Iasi, Romania, October 2014.

[13] R. Deng, R. Lu, C. Lai, and T. H. Luan, "Towards power consumption-delay tradeoff by workload allocation in cloud-fog computing," in *Proceedings of the IEEE International Conference on Communications (ICC)*, pp. 3909–3914, London, UK, June 2015.

[14] A. A. Mutlag, M. K. Abd Ghani, N. Arunkumar, M. A. Mohammed, and O. Mohd, "Enabling technologies for fog computing in healthcare IoT systems," *Future Generation Computer Systems*, vol. 90, no. 1, pp. 62–78, 2019.

[15] A. George, H. Dhanasekaran, J. P. Chittiappa, L. A. Challagundla, and S. S. Nikkam, "Internet of Things in health care using fog computing," in *Proceedings of the IEEE Long Island Systems, Applications and Technology Conference (LISAT)*, pp. 1–6, Farmingdale State College, USA, May 2018.

[16] K. S. Awaisi, S. Hussain, M. Ahmed, A. A. Khan, and G. Ahmed, "Leveraging IoT and fog computing in healthcare systems," *IEEE Internet of Things Magazine*, vol. 3, no. 2, pp. 52–56, 2020.

[17] S. Oueida, Y. Kotb, M. Aloqaily, Y. Jararweh, and T. Baker, "An edge computing based smart healthcare framework for resource management," *Sensors*, vol. 18, no. 12, p. 4307, 2018.

[18] A. P. dos Santos, D. W. S. Lima, F. S. Freitas, and G. M. da Silva, "A motivational study regarding IoT and middleware for health systems," in *Proceedings of the The Tenth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, pp. 102–106, Venice, Italy, October 2016.

[19] G. Sun, F. Yu, X. Lei, Y. Wang, and H. Hu, ""Research on mobile intelligent medical information system based on the internet of things technology," in *Proceedings of the 8th International Conference on Information Technology in Medicine and Education (ITME)*, pp. 260–266, Fuzhou, China, December 2016.

[20] K. Ullah, M. A. Shah, and S. Zhang, "Effective ways to use Internet of Things in the field of medical and smart health care," in *Proceedings of the International conference on intelligent systems engineering (ICISE)*, pp. 372–379, Islamabad, Pakistan, January 2016.

[21] M. Yannuzzi, R. Milito, R. Serral-Gracià, D. Montero, and M. Nemirovsky, "Key ingredients in an IoT recipe: Fog computing, cloud computing, and more fog computing," in *Proceedings of the IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pp. 325–329, Athens, Greece, December 2014.

[22] M. Ijaz, G. Li, L. Lin, O. Cheikhrouhou, H. Hamam, and A. Noor, "Integration and applications of fog computing and cloud computing based on the internet of things for Provision of healthcare services at home," *Electronics*, vol. 10, no. 9, p. 1077, 2021.

[23] B. Farahani, F. Firouzi, V. Chang, M. Badaroglu, N. Constant, and K. Mankodiya, "Towards fog-driven IoT eHealth: promises and challenges of IoT in medicine and healthcare," *Future Generation Computer Systems*, vol. 78, no. 2, pp. 659–676, 2018.

[24] P. Verma and S. K. Sood, "Fog assisted-IoT enabled patient health monitoring in smart homes," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1789–1796, 2018.

[25] N. Constant, D. Borthakur, M. Abtahi, H. Dubey, and K. Mankodiya, "Fog-assisted wiot: A smart fog gateway for end-to-end analytics in wearable internet of things," in *Proceedings of the The 23rd IEEE Symposium on High Performance Computer Architecture*, pp. 1–5, Austin, Texas, USA, February 2017.

[26] R. Al-Mashhadani, G. Alkawsi, Y. Baashar, A. Ahmed Alkahtani, and F. Hani Nordin, "Deep learning methods for solar fault detection and classification: A review," *Information Sciences Letters*, vol. 10, no. 2, p. 13, 2021.

[27] A. Al-Sammarraee and N. J. I. S. L. Alshareeda, "The role of artificial intelligence by using automatic accounting information system in supporting the quality of financial statement," *Information Sciences Letters*, vol. 10, no. 2, p. 8, 2021.

[28] I. Atia, M. L Salem, A. Elkholy, W. Elmashad, and G. J. I. S. L. Am Ali, "In-silico analysis of protein receptors contributing to SARS-COV-2 high Infectivity," *Information Sciences Letters*, vol. 10, no. 3, p. 19, 2021.

[29] Y. M. Al-Rawi, W. S. Al-Dayyeni, and I. J. I. S. L. Reda, "COVID-19 impact on education and work in the kingdom of bahrain: survey study," *Information Sciences Letters*, vol. 10, no. 3, p. 5, 2021.

[30] S. Sengan, O. I. Khalaf, D. K. Vidya Sagar, D. K. Sharma, Q. A. J. Prabhu, and A. A. Hamad, "Secured and privacy-based IDS for healthcare systems on E-medical data using machine learning approach," *International Journal of Reliable and Quality E-Healthcare*, vol. 11, no. 3, pp. 1–11, 2022.

[31] S. A. Syed, K. Sheela Sobana Rani, G. B. Mohammad, K. K. Chennam, and R. Jaikumar, "Design of resources allocation in 6G cybertwin technology using the fuzzy neuro model in healthcare systems," *Journal of Healthcare Engineering*, vol. 2022, 2022.

[32] M. E. Karar, M. F. Al-Rasheed, A. F. Al-Rasheed, and O. Reyad, "IoT and neural network-based water pumping control system for smart Irrigation," *Information Sciences Letters*, vol. 9, no. 2, pp. 107–112, 2021.

[33] S. M. E. I. Ammer, "Content analysis of lighting and color in the embodiment of fear concept in horror movies: a Semiotic

approach," *Information Sciences Letters*, vol. 9, no. 2, pp. 135–142, 2021.

[34] M. E. Karar, F. Alotaibi, A. A. Rasheed, and O. Reyad, "A pilot study of smart agricultural irrigation using unmanned aerial vehicles and IoT-based cloud system," *Information Sciences Letters*, vol. 10, no. 1, pp. 131–140, 2021.

Retraction

Retracted: Application Effect Analysis of Operating Room Detailed Nursing Based on Medical Big Data in Patients Undergoing Gastrointestinal Tumor Surgery

Journal of Healthcare Engineering

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

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

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

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

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

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

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

References

[1] Y. Meng, A. Sun, G. Ji, C. Wei, and J. Jia, "Application Effect Analysis of Operating Room Detailed Nursing Based on Medical Big Data in Patients Undergoing Gastrointestinal Tumor Surgery," *Journal of Healthcare Engineering*, vol. 2022, Article ID 8575305, 14 pages, 2022.

Research Article

Application Effect Analysis of Operating Room Detailed Nursing Based on Medical Big Data in Patients Undergoing Gastrointestinal Tumor Surgery

Yan Meng [ID],[1] Aixue Sun,[2] Ge Ji,[3] Caiye Wei,[1] and Junhong Jia[1]

[1]*Operating Room, Xingtai People's Hospital, Xingtai 054000, Hebei, China*
[2]*Oncology Department, Xingtai People's Hospital, Xingtai 054000, Hebei, China*
[3]*Operating Room Recovery Room, Xingtai People's Hospital, Xingtai 054000, Hebei, China*

Correspondence should be addressed to Yan Meng; 201901224238@stu.zjsru.edu.cn

With the continuous development of internet information computing, the continuous improvement of medical and health systems, and the continuous increase of medical big data, traditional operating room care also needs to be further optimized. Medical big data is a forum data set for medical industry healthcare, electronic medical record information, clinical case record information, medical financial data, remote patient monitoring data, clinical decision support data, medical insurance data set, online consulting platform, and so on. Gastrointestinal tumors are currently one of the largest malignant tumors. Compared with ordinary patients, the presence of fear, depression, irritability, and other unhealthy emotions in patients with gastrointestinal tumors will reduce the therapeutic effect. Without careful care, the use of chemotherapy and other treatments makes patients vulnerable to various side effects. This article aims to study the use of medical big data intelligent algorithms to perform detailed care for patients during gastrointestinal tumor surgery and analyze the effects of care. This paper proposes an improved DNN algorithm; the DNN algorithm is to use several weight coefficient matrices and bias vectors to perform a series of linear operations and activation operations with the input value vector, starting from the input layer, backward calculation layer by layer, until the operation reaches the output layer, and the output result is obtained. This algorithm is used to study the theory, use mathematical formulas for method calculation and model design, and use the model to carry out detailed nursing experiments in the relevant operating room. The results of the experiment show that patients who have performed detailed care have a 27.2% improvement in treatment and rehabilitation effects than those who have not, and the level of detailed care has an obvious positive relationship with the rate of condition conversion. In the end, the hospital's detailed care quality evaluation index, which is QEI, increases by 1 point, which can increase the rate of condition conversion by 0.4.

1. Introduction

1.1. Background and Significance. In recent years, internet medical care has gradually been accepted and widely used by more and more people. Internet medical treatment optimizes medical procedures and alleviates problems such as uneven distribution of medical resources and asymmetry of medical information. Nowadays, the population's subhealth status is getting worse, and the incidence of chronic diseases and senile diseases is increasing. Traditional medicine uses doctors to diagnose and treat diseases. At present, with the

rapid development of medical information technology, disease prevention is more important than disease diagnosis and treatment. As an important part of the hospital, the operating room is an important place for the treatment of diseases. The effectiveness and safety of surgical treatment depend not only on the doctor's surgical skills but also on the quality and level of nursing work. The application of detailed nursing in the operating room can effectively reduce the impact of anesthetics and surgery on the patient, relieve the psychological pressure of the patient, improve the success rate of the patient's surgery, and provide a guarantee for the

recovery of health after the operation. Detailed nursing is a new type of superior nursing model. It requires nurses in the operating room to have proficient professional qualities and strong psychological qualities and at the same time pay attention to the details of the work, minimize the risks, and ensure the safety of the patients during the hospitalization as much as possible.

Gastrointestinal cancer is a common malignant tumor. However, because of the special location of the disease, many patients may be reluctant to receive treatment. As time delays, the tumor will grow larger and eventually reach the point where it cannot be treated. Therefore, the care of patients with gastrointestinal tumors is a kind of meticulous treatment. To achieve a more ideal treatment effect, treatments need to be carried out from various aspects such as psychology, diet, and daily life, so as to improve the quality of life of the patients and better adhere to the treatment. Medical big data uses the Internet to conduct online medical consultations for patients, understand the physical and mental conditions of patients at any time, inquire and communicate online, encourage patients to receive treatment in a timely manner, and be more intelligent and humane. The medical big data establishes the intelligent operating room, the establishment of an intelligent operating room, that is, the need to establish a digital operating room, intelligent drug management, intelligent consumables management, operating room personnel behavior management, intelligent logistics robot, intelligent hardware terminal, and so on. Intelligent hardware terminal includes intelligent logistics management system, digital integrated operating room system, intelligent drug management system, intelligent consumable management system, and operating room personnel behavior management system. Through joint development with hospitals, deep research in different types of hospitals such as cardiovascular and tumor, we constantly understand user needs and provide users with modular configuration solutions for different hospital levels and types, so as to achieve the perfect solution for user management needs, which can monitor the patient's physical condition in real time during the operation. Medical big data analysis provides medical care for the public, raises the threshold of disease prevention, and truly simplifies precision medicine, medical care, and public health.

1.2. Related Work. Some scholars have carried out relevant studies on the detailed care of patients with gastrointestinal tumor surgery. Gastrointestinal stromal tumors (GIST) account for 0.1–3% of gastrointestinal malignancies. Surgery is the main method of treatment, but in high-risk tumors, imatinib (tyrosine kinase inhibitor is a small molecule protein kinase inhibitor, which has the effect of blocking one or more protein kinases. It is clinically used to treat chronic myelogenous leukemia and malignant gastrointestinal stromal tumors) helps obtain better oncology results. Navarrete et al. introduces a rare gastric GIST patient. Despite the neoadjuvant treatment of imatinib, it has undergone a very thorough evolution in a short period of time [1]. A malignant gastrointestinal neuroectodermal tumor (GNET) is a rare soft tissue sarcoma and a unique clinical pathological entity recently described. So far, only a few cases have been reported in the literature, so little is known about the behavior and diagnosis of the tumor. A case of the gastrointestinal neuroectodermal tumor with liver metastasis studied by Keditsu et al. can effectively inhibit the tumor; treatment of metastatic tumors in the liver is often ineffective. Systemic chemotherapy can temporarily shrink the mass and prolong life but is highly dependent on the primary site and cannot cure the disease. Some medical centers recommend hepatic arterial catheterization in selected cases, which has a significant effect on the treatment of most liver diseases [2]. The functional significance of missing microRNA has been reported in several human malignancies. An abnormally expressed transmembrane glycoprotein can effectively inhibit tumor growth in mice. However, the lack of an effective tumor-specific delivery system remains an unsolved clinical challenge for the successful translation of microRNA. Setua et al. has developed a miRNA-145-based magnetic nanoparticle formulation (miR-145-MNPF) and evaluated its anticancer efficacy. The expression of miR-145 (miRNA-145) plays an important regulatory role in the growth inhibition, metastasis, invasion, and chemosensitivity of various tumors. The high expression of miRNA-145 in tumor tissue cells can inhibit the activation of proto-oncogenes, while the downregulation of miRNA-145 expression can reduce the ability to inhibit the expression of proto-oncogenes and the growth, proliferation, and even metastasis of cancer cells. Experimental results show that miR-145-MNPF shows the best particle size and zeta potential, can effectively internalize and restore miR-145 in pancreatic cancer cells, and inhibit cell proliferation, clone formation, migration, and invasion of pancreatic cancer cells [3]. A patient's examination revealed a significantly enlarged gastrointestinal mass (2 cm in diameter) located in the gastric antrum near the pylorus at the greater curvature. EUS-FNA was performed with the assistance of on-site cytopathology to assess the adequacy of the material. The cytopathological smear showed a uniform group of round epithelioid cells with relatively small nucleoli and variable eosinophilic cytoplasm, stained with smooth muscle actin, but desmin, chromogranin, synapse, and keratin were negative, and the final cytological diagnosis of gastric globular tumor (GT) was obtained. The patient underwent a CT scan, which confirmed the presence of high-density lesions of approximately 14 mm and no lymphadenopathy or metastatic disease. There were no contraindications to surgery, so Castro Ruiz et al. performed gastric laparoscopic wedge resection (it is commonly used for surgical removal of lesions on the eyelids, nostrils, auricles, and lips. This kind of excision involves the full thickness and edge of local tissues or organs. The excised lesions are wide on the surface and narrow in the deep, in the shape of a wooden wedge. The sutures should also be sutured in layers) of the tumor. Laparoscopy is a very good minimally invasive technique for small and benign tumors. The patient is discharged on the fourth day after surgery. Macroscopically, the lesion has clear borders and is rubber-like. Histological examination confirmed the cytological diagnosis of gastric GT, and the

surgical margin was disease-free [4]. Although treatment with imatinib in patients with advanced gastrointestinal stromal tumors (GIST) has brought significant clinical benefits, the disease will eventually progress due to resistance to imatinib. Treatment options after the failure of first-line imatinib include dose escalation of imatinib or switching to sunitinib. However, there is currently no large-scale study to compare the differences in the efficacy of these two treatment strategies or the effects of surgery. Hsu et al. and other studies recruited hundreds of patients with advanced GIST, and surgery observed a favorable survival trend, although this advantage disappeared after adjusting for confounding factors. Hsu et al. were divided into three groups for patients with advanced GIST who no longer responded to first-line imatinib. The treatment success rate and surgical effect of the treatment plan were evaluated by the overall survival rate. It is concluded that for patients with advanced GIST who have failed first-line imatinib treatment, compared with directly switching to sunitinib, the dose escalation of imatinib can significantly prolong OS [5]. A woman in her 40s initially presented with anal pain and was diagnosed with rectal GIST. A 9 cm tumor extends near the anus and requires radical resection of the abdominal and perineal tumor. The patient initially received neoadjuvant treatment with imatinib mesylate. Six months of neoadjuvant chemotherapy reduced the tumor to about 47% of its original size and allowed anal-sparing surgery. Imatinib mesylate can inhibit Bcr-Abl tyrosine kinase at the cellular level in vitro and in vivo and can selectively inhibit Bcr-Abl positive cell line cells, Ph chromosome positive chronic myeloid leukemia, and acute lymphoblastic leukemia proliferation and induction of apoptosis in fresh cells from patients. In addition, imatinib mesylate can also inhibit platelet-derived growth factor (PDGF) receptor; stem cell factor (SCF); c-Kit receptor tyrosine kinase, thereby inhibiting PDGF; and stem cell factor-mediated cellular behavior. Research of Kenno et al. showed that from the perspective of preserving anal function, the neoadjuvant treatment of imatinib mesylate is useful for large rectal GIST [6]. Because the tumor is located in the pelvis, anal sparing surgery for large rectal gastrointestinal stromal tumors (GIST) can be difficult. Therefore, rectal GIST may require extensive surgery, such as abdominal perineal resection. In recent years, in some cases, preoperative imatinib treatment has been used to reduce tumor size and protect the anus. However, there are few reports of laparoscopic anus-sparing surgery for giant rectal GIST. In a 55-year-old male case, a 10 cm pelvic mass was found in the lower rectum. Sn A et al. performed ultrasound endoscopy with fine-needle aspiration, and the pathological results led to the diagnosis of GIST. The mass has spread to the prostate and left levator ani muscle, so surgery is considered difficult without damaging the pseudocapsule. Therefore, imatinib mesylate (IM) was used for 8 months of chemotherapy before surgery; the mass was reduced to 7.8 cm; and laparoscopic intersphincterectomy was performed. Among them, the patient underwent IM chemotherapy before surgery and then performed anus preservation surgery. In a review of patients who successfully underwent anal-sparing laparoscopic surgery after preoperative IM chemotherapy, this case represented the largest tumor size. It is concluded that preoperative imatinib mesylate chemotherapy can effectively reduce rectal GIST, and laparoscopic ISR can be used to preserve the anus, even if the tumor is large [7].

1.3. Innovation. (1) This article improved the model of knowledge map combined with deep learning to realize disease self-diagnosis; (2) the ranking learning is applied to doctor recommendation, and the characteristics of doctors are comprehensively considered for ranking; (3) this article enriched the research angles and research methods of the doctor-patient matching problem; and (4) this article diversified detailed care that made the treatment more efficient and the patients feel more at ease.

2. Detailed Nursing Plan under Big Data

2.1. Medical Big Data. With the rapid development of medical information engineering, the complex types and scale of medical data make the existing medical big data analysis technology unable to meet its needs. It is difficult to collect, analyze, and process medical data in a timely and streamlined manner and integrate it into useful information to support accurate medical decision-making. Therefore, how to efficiently calculate and integrate medical big data becomes more and more important [8].

2.1.1. Sources of Medical Big Data. Medical big data mainly comes from clinical medicine, biomedicine research and development, new medical models, and information-based treatment. The detailed source distribution can be shown in Figure 1.

2.1.2. The Characteristics of Medical Big Data. (1) Multiple data types include photos, audio, video, Bluetooth information, sensors, signal search, and so on. (2) Data sources are wide, such as the four main sources mentioned above. (3) A large amount of data needs to be stored. CT images contain about 100 MB of data, and the size of each genome sequence file is about 750 M. The high-resolution standard pathology map contains approximately 5 GB of medical data information. The unit of measurement for big data is at least a scientific measurement unit with 100 million as the unit. (4) Data development and update are fast; research on medical convenience is constantly updated and disseminated; and data change rapidly, which is different from traditional data mining technology. (There are mainly clustering methods, analysis methods, and regression methods. Traditional data mining techniques are all based on the development of a centralized underlying software architecture, which is difficult to parallelize and has low execution efficiency.) (5) Business value is great; health is very important; and big data analysis is used to understand the development mode and development trend of things and obtain valuable business information [9].
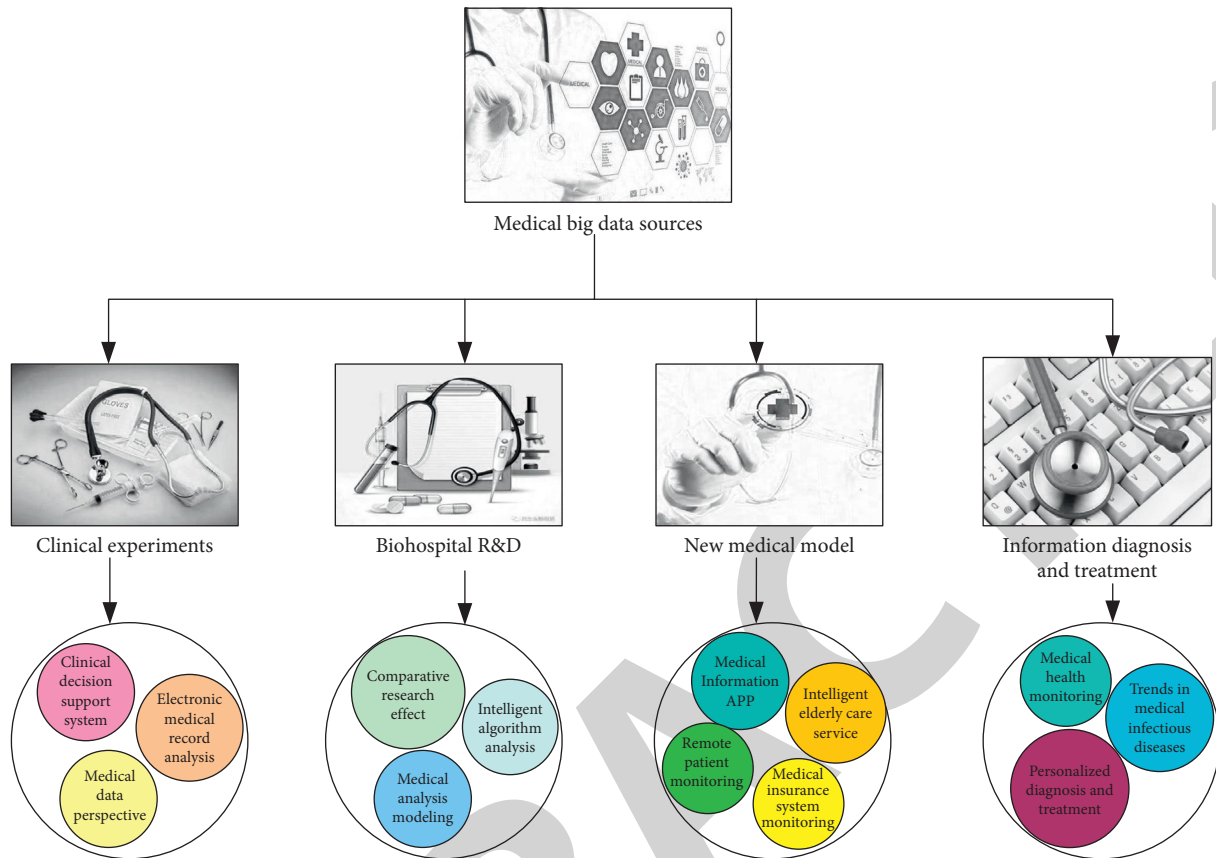
Figure 1: Sources of medical big data.

2.1.3. Application Scenarios of Medical Big Data. The current important application scenarios are: (1) Clinical decision support system. The data-driven clinical decision support system can make use of big data analysis technology to make itself more intelligent and improve the work efficiency of medical staff and the quality of medical services. (2) Remote monitoring through various wearable health equipment, remote monitoring of patients with chronic diseases, and remote recording of relevant data. Through the collection and analysis of big data, it can help healthcare professionals develop treatments for patients. (3) Comprehensively comparing and analyzing the patient's individual characteristics, disease-related data, and treatment effect data; thoroughly comparing various treatment methods; and finally determining the best treatment plan for a specific patient and conducting effective research. Big data visualization improves the transparency of medical data and processes, promotes the optimization of medical business processes, and reduces medical costs while improving the quality of medical services. And it regulates the medical behaviors between doctors and patients, makes medical behaviors more transparent and effective, reduces medical inconsistencies, and reduces medical disputes. Since the development of medical big data in 2005, the data have grown from several thousand T to hundreds of millions of T; and the application of data has grown exponentially. According to the collected statistics, it can be shown in Figure 2 [10].

2.1.4. Medical Big Data Algorithm. At present, there are many kinds of big data algorithms, which can be described as a hundred flowers blooming. For the common algorithms of medical big data, take the neural network algorithm as an example. Although traditional cloud computing can perform a large amount of data analysis and basically meet the requirements of medical quality evaluation, there are still many shortcomings, such as a long calculation process and a large number of sample training and testing. The DNN intelligent cloud algorithm proposed in this paper can significantly reduce the calculation time, and the quality evaluation index QEI obtained can more accurately reflect the patient's service evaluation of the detailed surgical care and can be used to detect the hospital's treatment effect [11].

2.2. DNN Algorithm Model. This article classifies the QEI of each hospital into categories I, II, III, and IV; QEI between 0 and 25 is type I; QEI between 25 and 50 is type II; QEI between 50 and 75 is type III; and QEI between 75 and 100 is type IV. The larger the QEI, the better the quality of detailed care in the hospital and the higher the treatment effect. From this analysis of the model, the following formula can be obtained:

$$QEI = \lambda_1 \times PI + \lambda_2 DI, \tag{1}$$

where PI is the proximity index, DI is the disease index, and λ_1 and λ_2 are the proportional coefficients of PI and DI,
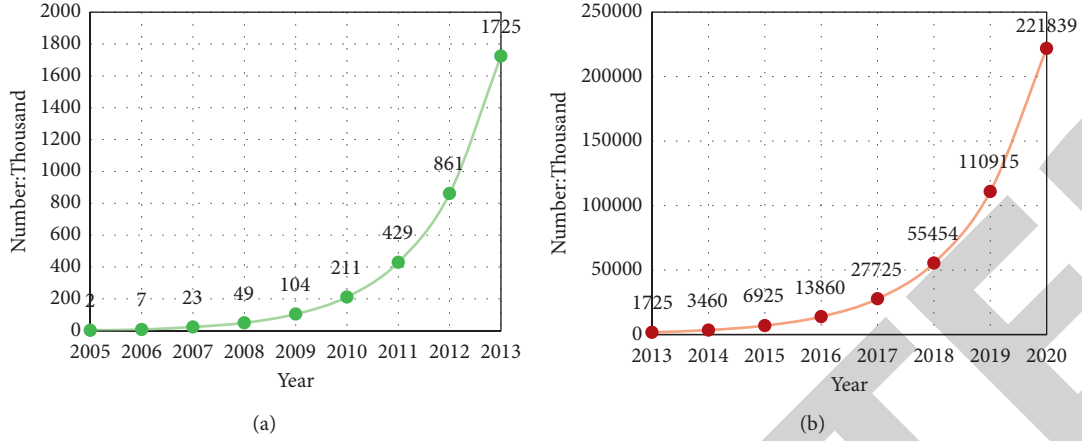
Figure 2: The growth of medical big data from 2005 to 2020.

respectively. It is related to the comprehensive medical level and medical equipment condition of each hospital. In order to get the actual value conveniently, this paper assigns values to PI and DI and λ_1 and λ_2, respectively, as shown in Table 1 [12].

Di is the disease-to-good rate, which can be obtained by formula conversion with the quality data value of the case known, and the conversion is shown in the following formula:

$$D_i = \frac{G + B}{S} \times 100\%, \quad (2)$$

where G is the number of optimistic conditions, B is the number of pessimistic conditions, and S is the number of total conditions.

P_i is the neighboring ratio of a hospital and can be calculated as follows:

$$P_i = \frac{m_i}{N_i}, \quad (3)$$

where m_i is the number of neighbors of the hospital and N_i is the total number of *data* of the hospital.

The neighboring point algorithm based on DNN can be understood as the neighboring ratio of the data x that needs to be detected is the distance from the x point to its i-th neighboring data. Formula (3) can be expressed as the area A where the data x is within its diameter:

$$A(x, r) = \{x \in D | r(x, i) \le R\}, \quad (4)$$

where r is the radius range value from point x to its i-th neighboring data, R is the maximum value, and D is the set of data x.

Therefore, the distance from point x to point i can be shown in formula (4):

$$L_i(x) = \{x \in D | r(x, i) \le l - r(i), x \neq i\}, \quad (5)$$

where l represents the number of neighbors.

Given D and l, then the l proximity distance of data x is shown in the following formula:

Table 1: Corresponding assignment for each indicator and parameter.

QEI genre	I	II	III	IV
PI	40	30	20	10
DI	10	20	30	40
λ_1	0.25	0.25	0.5	0.5
λ_2	0.4	0.4	0.6	0.6

$$D^l(i) = \frac{\sum_{x \in D} r(x, i)}{|Li|}. \quad (6)$$

This article has improved the previous algorithm. Assuming that the area range of the distance calculation between the two data is O, then

$$O = \left[\frac{N_i}{N_h}\right], \quad (7)$$

where N_h is the number set of the hospital. The purpose of this calculation is to make the number of data points be compared as much as possible to all the data points within the scope of the hospital [13]. Letting o be the midpoint data point of the class, and the number of data points in it should be at least n, take $d/2$ as the radius of the circle, and o as the center of the circle. For any two points, suppose x_1, x_2, then satisfy

$$r(x_1, x_2) \le \frac{d}{2}, \quad (8)$$

$$|N(x_1, x_2)| \ge n.$$

Filtering principle: When T has been assigned to the smallest value among n candidate neighboring points if it satisfies

$$D^l(i) + r(x_1, x_2) > T. \quad (9)$$

At this time, x_2 cannot be a neighboring point.

As shown in Figure 3, from x_1 and x_2 and the three neighboring points ω_1, ω_2, and ω_3, using the theorem that
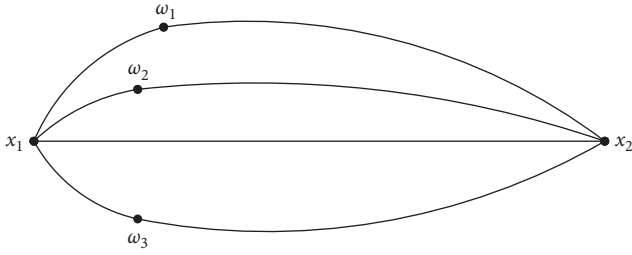
FigURE 3: Schematic diagram of the neighbors of x_1 and x_2.

the sum of the two sides is greater than the third side, we can know

$$
\begin{aligned}
r(x_1, \omega_1) + r(x_1, x_2) &> r(x_2, \omega_1), \\
r(x_1, \omega_2) + r(x_1, x_2) &> r(x_2, \omega_2), \\
r(x_1, \omega_3) + r(x_1, x_2) &> r(x_2, \omega_3).
\end{aligned}
\tag{10}
$$

It can be generalized to get the following formula:

$$
\sum_{i=1}^{n} {}_{\omega \in N(x_1, x_2)} r(x_1, \omega_i) + n \times r(x_1, x_2) > \sum_{i=1}^{n} {}_{\omega \in N(x_1, x_2)} r(x_2, \omega_i). \tag{11}
$$

That is, if both sides are divided by n, we can get

$$
\frac{1}{n} \sum_{i=1}^{n} {}_{\omega \in N(x_1, x_2)} r(x_1, \omega_i) + r(x_1, x_2) > \frac{1}{n} \sum_{i=1}^{n} {}_{\omega \in N(x_1, x_2)} r(x_2, \omega_i). \tag{12}
$$

That is:

$$
D^l(x_1) + r(x_1, x_2) > D^l(x_2). \tag{13}
$$

Combining formula (9), we know

$$
D^l(x_1) + r(x_1, x_2) < T. \tag{14}
$$

Therefore,

$$
D^l(x_2) < T. \tag{15}
$$

The above calculation can show that the neighboring point D^l is already greater than $D^l(x_2)$, and the neighbors of x_2 need to be filtered out.

Considering that the whole area can be divided into countless small areas, cluster analysis can be combined, assuming $\overline{O}$ is the average value of the data center points in a certain area. This area includes m data, then

$$
\overline{O}_i = \frac{\sum_{i=1}^{m} x_i}{m} = \frac{x_1 + x_2 + \cdots + x_m}{m}. \tag{16}
$$

The complexity of this clustering can be expressed by the following formula:

$$
\sigma = \frac{m}{n}(A + O) \times \left(m - \frac{(i-1)m}{x} \right). \tag{17}
$$

Then for the theoretically generated y cluster centers, their cluster center values are $\overline{O}_1, \overline{O}_2, \ldots, \overline{O}_y$, where $O_{\max}$ and $O_{\min}$, respectively, represent the maximum and minimum values in each cluster; then the assignment of each corresponding cluster center is

$$
Y(x) = \frac{\lfloor 10 \times (\gamma + ((\overline{O}_i - \overline{O}_1)/(\overline{O}_y - \overline{O}_1))) \rfloor}{10} = \frac{\lfloor 10 \times (\gamma + ((\overline{O}_i - O_{\min})/(O_{\max} - O_{\min}))) \rfloor}{10}, \tag{18}
$$

where $\gamma = 0.1$ in the formula is to keep one decimal place for easy calculation [14].

2.3. Detailed Care in the Operating Room.

Compared with traditional nursing work, the meticulous care in the operating room allows patients to feel more considerate service, let them carefully immerse themselves in the warm treatment environment, follow the hospital's arrangements, improve patient satisfaction, treat patients, provide convenience for the operation, smooth implementation, improve the quality of postoperative rehabilitation of patients, and increase patient satisfaction with nursing work [14].

2.3.1. The Importance of Detailed Care in the Operating Room.
(1) Adapting a variety of nursing models to the needs of patients. Many patients have preoperative anxiety, which affects their effectiveness in treating certain diseases and makes it difficult for them to recover. Under the meticulous nursing mode, medical staff answer patients' questions, effectively reduce surgical anxiety, and cooperate more

actively with treatment, which improves the success rate of surgery to a certain extent. (2) Reducing postoperative complications (the statistical analysis method of classifying the clustered objects according to the characteristics of the things themselves; its purpose is to divide the data set according to a certain similarity measure) and improving patient care satisfaction. The medical staff will promptly inform the patient of the precautions after the operation, which can effectively calm down the patient's negative emotions. During the operation, the patient's physical characteristics can be detected and tracked in time. When a bad phenomenon occurs, it can be reported to the doctor in time, which can improve the treatment effect of the patient. At the same time, it also establishes a good image in front of the family and improves patients' satisfaction with long-term care. (3) Improving patient satisfaction and improving hospital reputation. In the advanced nursing mode, the nurse will explain the surgical precautions to the patient and family members when making an appointment for surgery to treat the disease. Family members have a deep understanding of the risks of treatment and the impact of surgery

on patients. Medical staff will do their best to improve patients to a certain extent and increase satisfaction and hospital reputation. Figure 4 shows the conventional process of detailed care in the operating room [15].

2.3.2. Application Measures of Detailed Nursing. (1) Preoperative care. It includes the following: following the doctor's surgical plan to ensure that relevant equipment and items are installed on time and disinfection and sterilization are handled properly, developing detailed nursing procedures in the operating room, clarifying the job responsibilities of all levels of personnel related to surgery, classifying nursing tasks at different times, assigning nursing tasks to specific nursing staff, visiting patients in the ward, reading cases, communicating with patients and their families, observing patients' conditions, explaining surgical precautions, helping patients overcome tension, solving patient-related disease knowledge, stabilizing patients, and improving patient compliance. (2) Intraoperative care. Nursing staff maintain easy communication with patients, making patients more reliable. During the anesthesia process, the nursing staff can comfort and encourage the patient with gentle words, touch the patient's body gently, and improve the patient's safety awareness. The patient inspection system should be observed during the operation, and the nurses should adopt a standardized and gentle attitude to avoid infringing on the privacy of patients. During the operation, proper insulation care and proper insulation of the exposed area are carried out to reduce the discomfort of the patient. During surgery, the liquid is used to maintain the room temperature and improve patient comfort. Carefully observe and record the patient's vital signs and minimize the entry and exit of nonsurgeons to prevent pathogens from entering the operating room. During the operation, ensure that the functions of various instruments are normal and the surgical materials meet the requirements of the specification and are completely exposed to the doctor's field of vision to ensure the smooth progress of the operation. (3) After observing the patient's physical characteristics, the patient was fully awake; the side effects disappeared; and he was sent to the ward for rehabilitation and handed over work with the ward nurse. After the operation, the nurse went to the ward to understand the wound healing of the patient in time, help the patient restore function, explain the precautions, and request the evaluation of the patient's care quality [16, 17].

3. The Detailed Nursing Experiment of Gastrointestinal Tumor Based on Big Data

3.1. Design of Medical Big Data Model

3.1.1. Data Collection [18]. The big data mentioned in this article is geospatial big data, which mainly includes POI data and pedestrian navigation data. Among them, POI data is also called point of interest, which generally refers to point data on internet electronic maps. This includes vector point data with four main attributes: name, address, coordinates, and category. Pedestrian navigation data refers to route planning data between multiple starting points and multiple destinations obtained through the electronic map API interface, which mainly includes route planning distance and driving time. The data terminal can be divided into multiple types, as shown in Figure 5 [19, 20].

This article investigated 29 hospitals for relevant data collection. The information was collected based on the comprehensive level of the hospital Q, the hospital's neighboring area A, the number of monthly illnesses S, the medical history of gastrointestinal tumors n, and the rate of disease conversion D_i, and so on, and the results are shown in Table 2.

3.1.2. Data Processing. (1) Data preprocessing. JSON is a lightweight data exchange format, and its concise and clear hierarchical structure makes it an ideal data exchange language. In order to facilitate data processing, this research saves POI data and walking time data in the form of json format and data frame objects. With the help of Pandas' powerful data aggregation capabilities, we can splice, transform, sort, and aggregate the acquired POI and walking time data to provide a more easily processed data format for data vectorization. (2) Data conversion. Different data sources are directly connected to the different coordinate systems in the above vector file. Therefore, before performing spatial statistical analysis and calculations, vector files must be coordinated and corrected to convert them into computer-recognizable data. (3) Establishing a spatial database. Add attribute data corresponding to point, line, and area elements for each administrative division, such as residential area label, resident health facility level, population, and so on. On this basis, establish a geographic database; import residential areas, medical facilities, administrative divisions, roads, and other elements into the database; and build the foundation for subsequent data processing, analysis, and visualization [21, 22].

This article sorts out the data in Table 2 and analyzes the relationship between A and S. The relationship between Q and D_i can be obtained as shown in Figure 6.

It can be seen from Figure 6(a) that A and S are roughly proportional, that is, the larger the neighboring point area, the more the number of nearby doctors. Although A is very small in some areas, the number of illnesses S is still very large, such as the number 15 hospital. The reason can be seen in Figure 6(b). Although the comprehensive level of No. 15 hospital is relatively low, it has a high rate of disease-turning yield. From the overall relationship diagram, there is no absolute proportional relationship between the hospital's comprehensive level Q and the disease-turning yield rate D_i [23]. In order to analyze the reasons, this article conducted a detailed nursing management survey on several hospitals with the same characteristics and found that these hospitals do better in detailed nursing than other hospitals.

In order to further find the relationship between detailed nursing and its rate of disease conversion, this article excludes the interference of other factors and compares and analyzes the specialized gastrointestinal tumor hospitals numbered 3, 4, 7, and 9. Assuming that the level of detailed

```
┌─────────────────────┐       ┌──────────────────────────┐
│  Preoperative visit │──────▶│ On the day before the    │
└─────────────────────┘       │ operation, the roving    │
          │                   │ nurses used written and  │
          ▼                   │ oral forms to provide    │
┌──────────────────┐  ┌───────│ psychological care for   │
│ Fill in the      │◀─│ Go to │ the patient, and         │
│ patient handover │  │ the   │ explained the relevant   │
│ form and check   │  │ ward  │ precautions to relieve   │
│ the patient      │  │ to    │ the patient's worries.   │
│ information       │  │ pick  └──────────────────────────┘
└──────────────────┘  │ up... │
                      └───────┘
```
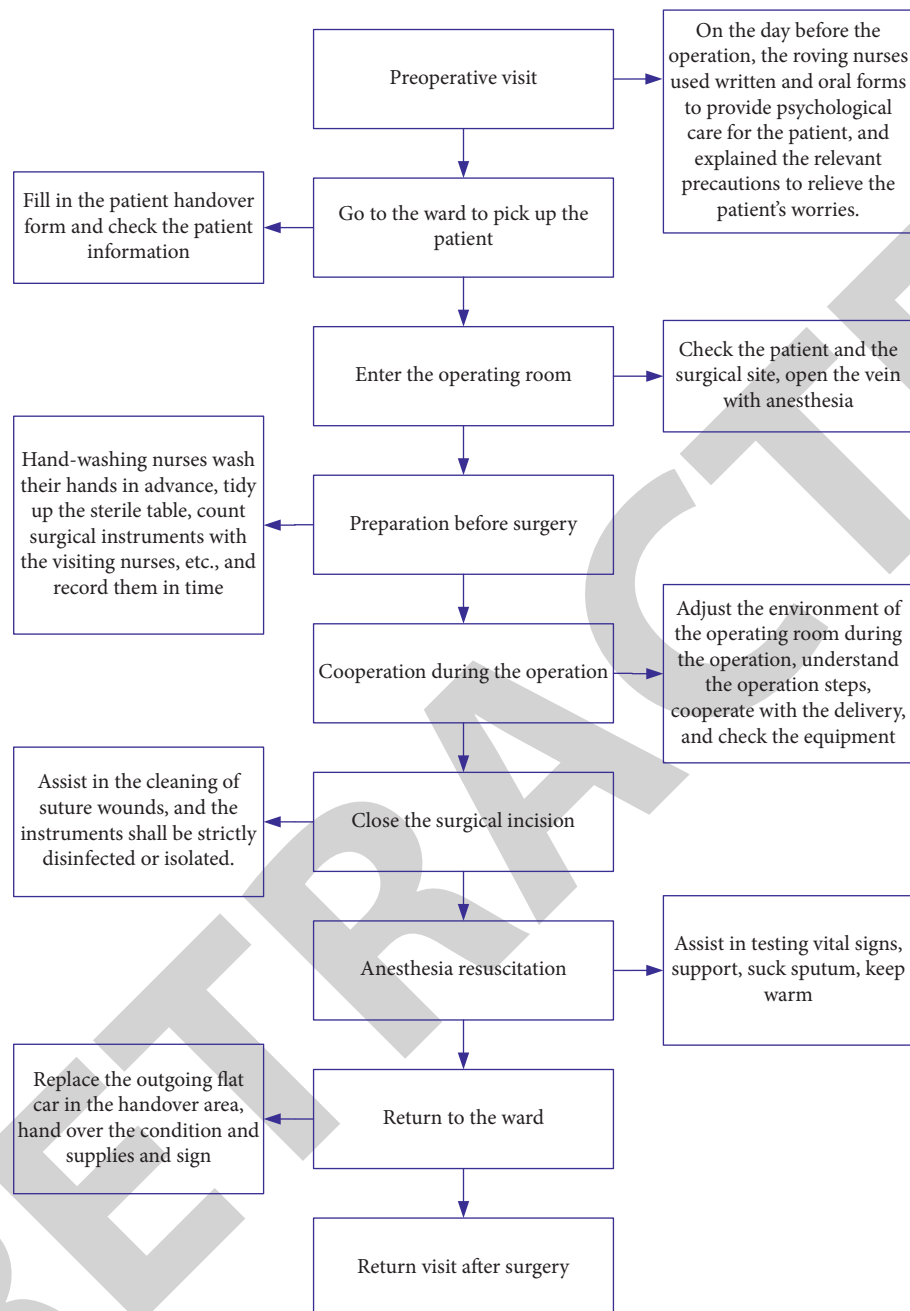
Figure 4: Detailed nursing process in the operating room.

nursing management is I, it can be divided into four levels: I, II, III, and IV, the higher the level, the more refined the care. The collated data is shown in Table 3:

After comparing the data in Table 3, it is found that the relationship between Q, A, N, and D_i is not obvious. From Table 3, it can be seen that the comprehensive level of hospital nos. 4 and 7 is quite high, no. 4 is slightly higher, and the medical records of gastrointestinal tumors are also similar, but the rate of disease conversion of no. 4 is lower than that of no. 7. The reason is that the level of detailed care management is higher and the detailed care is better. Similarly, observe hospital nos. 3, 7, 4 9; although the comprehensive level of the two groups is different, because

the level of detailed management is similar, the rate of disease conversion is very close. There is an obvious proportional relationship between I and D_i. In order to specify the value to express the relationship, detailed nursing experiments are carried out below in this article [24, 25].

3.2. Detailed Nursing Experiment

3.2.1. Experimental Subjects. One hundred surgical patients are included if they meet the following criteria: (1) all signed the informed consent; (2) no major disease; (3) no endocrine
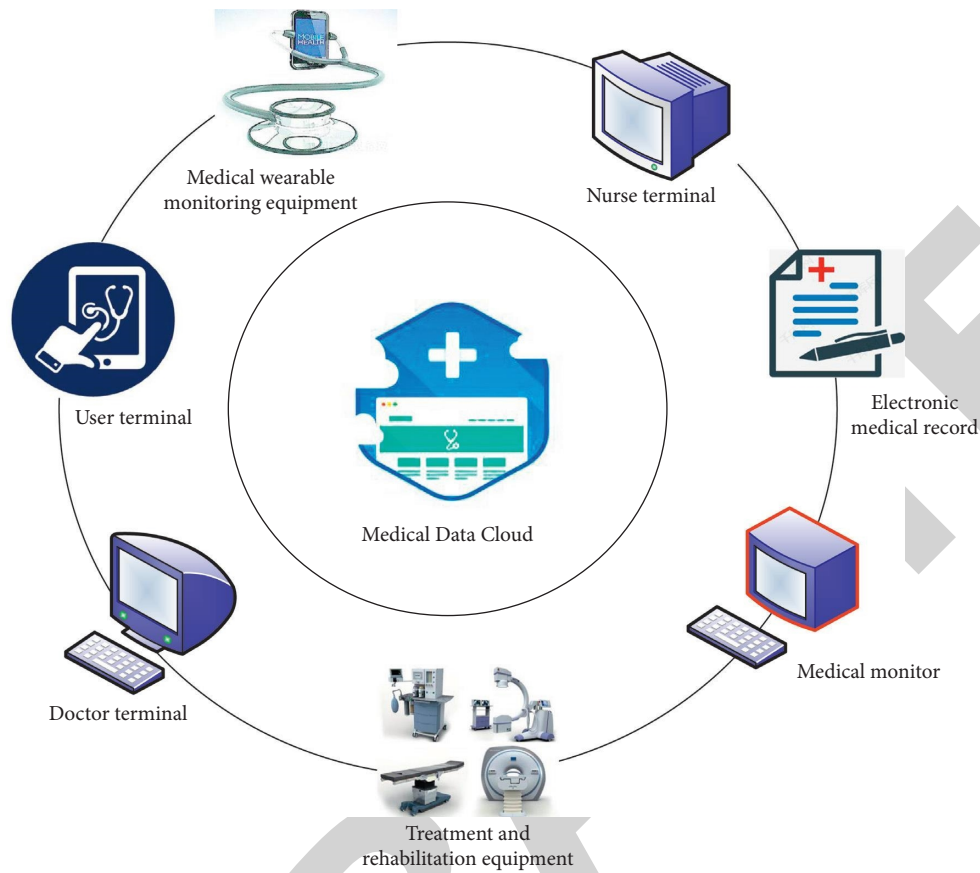
FIGURE 5: Data collection of the medical data cloud.

TABLE 2: Hospital-related data collection information.

Hospital code	Q	A	S	n	D_i
1	96	1,299	1,594	273	70
2	92	1,600	2,006	213	96
3	87	888	1,510	1,892	74
4	96	406	926	1,870	67
5	88	2,945	3,762	442	75
6	87	2,425	3,616	359	72
7	91	918	1,974	1,861	74
8	97	1,837	2,946	357	92
9	72	2,926	4,059	1,901	65
10	83	979	2,085	25	67
11	85	333	1,451	696	62
12	71	1,530	2,581	101	81
13	90	2,029	7,929	818	86
14	97	776	6,233	522	72
15	79	538	6,078	507	83
16	95	333	5,989	580	92
17	84	2,632	8,555	751	81
18	66	2,804	8,040	39	98
19	79	2,878	8,877	621	88
20	65	1,856	3,915	106	88
21	66	2,502	4,958	280	86
22	83	1,021	3,397	206	80
23	73	1,334	3,593	175	80
24	88	832	3,313	72	79
25	84	700	1,532	219	91
26	77	2,225	3,293	162	99
27	93	2,066	3,901	127	60
28	93	114	792	250	81
29	73	887	3,020	177	63

FIGURE 6: Relationship diagram of hospital-related data: (a) the relationship between the neighboring area and the number of monthly illnesses and (b) the relationship between the comprehensive level of the hospital and the rate of disease conversion.
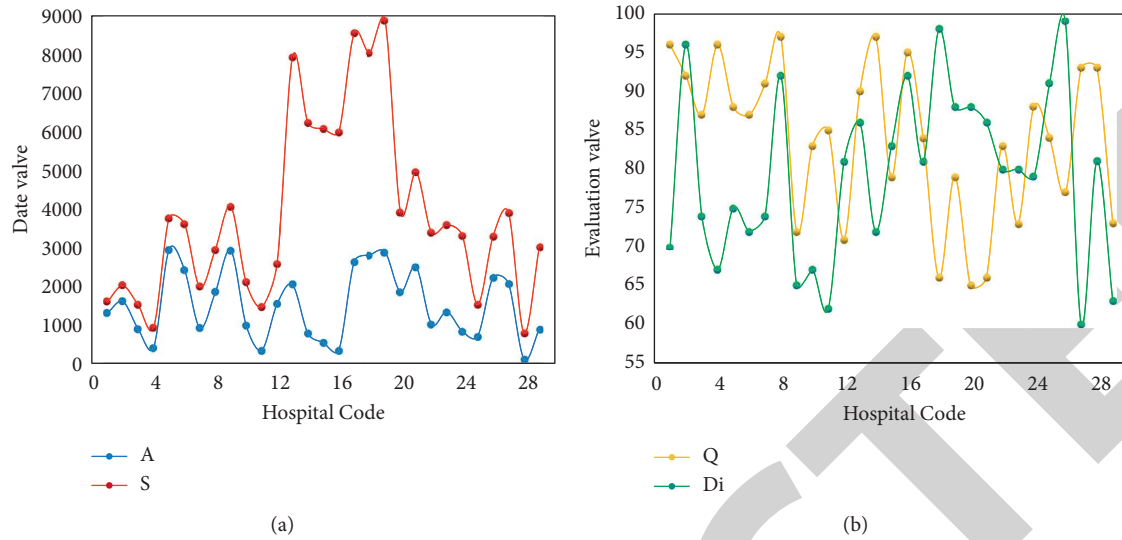
TABLE 3: Gastrointestinal cancer hospital data information comparison.

Hospital code	Q	A	N	D_i	I
3	87	888	1,892	74	II
4	96	406	1,870	67	I
7	91	918	1,861	74	II
9	72	2,926	1,901	65	I

or immune system disease before the operation; and (4) primary school education or above, which is enough to complete the investigation.

Excluding factors include: (1) patients with severe immune system diseases; (2) combined with diabetes, liver cirrhosis, cardiovascular infection, and lung infection; (3) people who have recently suffered from functional gastrointestinal disease and are taking gastrointestinal motility drugs; (4) patients with a history of mental and neurological diseases; and (5) patients with severe postoperative complications. Among the 100 selected patients, 41 were males, and 59 were females, aged 19 to 45 years old, and the operation time was 30 to 230 minutes. In a randomized controlled trial, patients were divided into an observation group and a control group, with 50 patients in each group. There was no statistically significant (i.e., the true degree of some statistical results of the data is credible, and there is a great grasp that the results of the data are not caused by chance) difference in general information between the two groups.

3.2.2. Experimental Method.

The control group adopts routine nursing management, and the experimental group applies refined nursing to nursing management. Nursing staff need to effectively manage the details and provide adequate care and love for patients involved in performing various clinical nursing tasks for cancer patients. Nursing staff can explain future nursing operations to cancer patients so that they can make proper psychological preparations before performing routine nursing operations. Compared with normal patients, the fear, depression, irritability, and other bad moods of patients with gastrointestinal stromal tumors reduce the therapeutic effect and are not conducive to the prognosis. Through chemotherapy and other treatments, patients are prone to various side effects. Nursing staff need to further strengthen communication, patiently understand the needs of patients, meet the reasonable requirements of patients, improve patients' treatment information, and prepare daily necessities related to patients. The specific steps are as follows: (1) convey information and education, (2) psychological detail care, (3) warm detail care, (4) meticulous care of intraoperative observation, and (5) detailed operation nursing [26, 27].

3.2.3. Evaluation Index.

Comparing nursing errors, inadequate nursing, nursing complaint rate, nursing quality, and postoperative recovery time, the nursing quality score is determined according to the hospital's self-made postoperative satisfaction survey score sheet. The total score is 100 points, which is directly proportional to the quality of care. Satisfaction with η is proportional to the quality evaluation index QEI, which can be converted by QEI = $k\eta$. Because according to the experiment, QEI is proportional to the disease, D_i, and the satisfaction is also proportional to D_i.

3.3. Experimental Results.

During the operation, the average communication times and the average communication
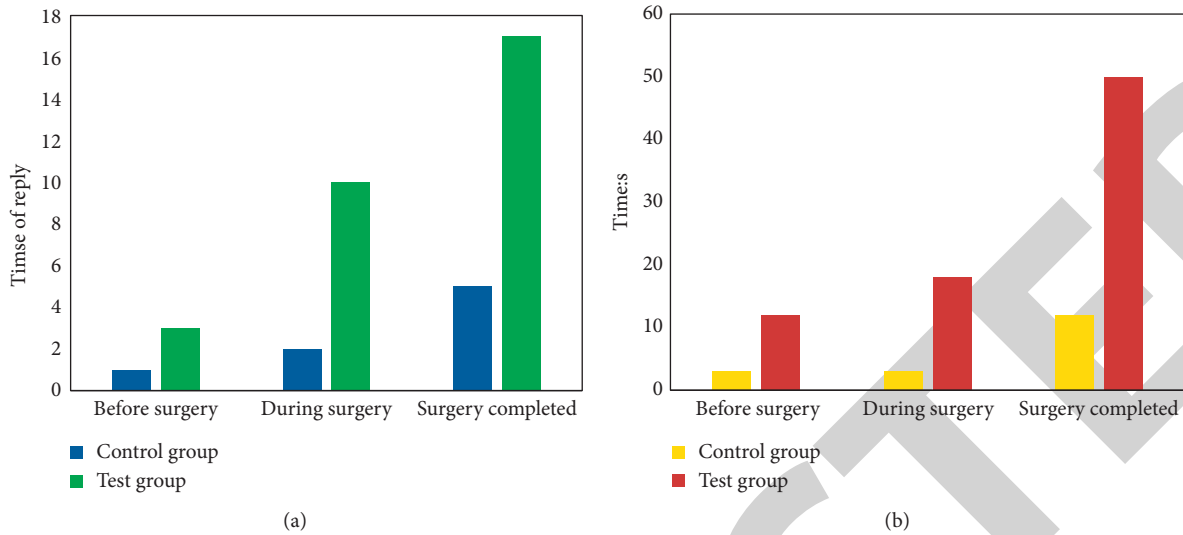
Figure 7: Comparison of patients' willingness to communicate: (a) comparison of patient response times and (b) comparison of patient communication time.

duration between the patient and the medical staff were compared, and the emotional value of the patient and the medical staff was compared as shown in Figure 7:

It can be seen from Figure 7 that for the experimental group that carried out detailed care, the number of patients' responses and the duration of communication were significantly greater, indicating that the patients' willingness to communicate was stronger. This is more conducive to solving the fears and doubts in the heart of the patient, calming the patient's psychology, and making the operation more smooth.

Then this article also compares the recovery of the two groups of patients after the operation, as shown in Figure 8.

It can be seen from Figure 8 that the average length of stay in the test group was 8 days, while of the control group was 11 days, and the recovery efficiency increased by 27.2%. And the test group that performed detailed care is shown in Figure 8(a), and the mood before the operation is relatively better. With the increase in the number of days in the hospital, the mood of the patients in the test group for detailed care has improved significantly, while in the control group, as shown in Figure 8(b), the mood of the patients rises relatively slowly.

After that, this article investigates the degree of detailed care and the satisfaction of the tested patients and their families with the hospital. The results obtained are shown in Table 4.

The satisfaction degree mentioned above is directly proportional to QEI. According to the different detailed care levels of test groups 1 and 2, the relationship between QEI and D_i can be obtained as shown in Figure 9.

It can be seen from Figure 9 that as the QEI value increases, the D_i value of the test group is also increasing, and the increase value has been very stable, while the D_i value of the control group is not stable with the increase of QEI. The increase in QEI is not through optimizing the management

level of detailed care but through other unstable factors. And the QEI value increases by 5 points; the D_i value increases by about 2 points, that is, the QEI increases by 1 point, which can increase the rate of disease conversion by 0.4 points. It can be concluded that the use of detailed care can actually and effectively improve the success of the surgery and the effect of treatment and rehabilitation for patients with gastrointestinal tumors [28].

In addition, in order to compare the big data processing performance of the DNN algorithm of this system with traditional algorithms, we run this system on the Hadoop cluster in this experiment to analyze medical big data. At the same time, we also need a single-node host with the same physical configuration for comparison. The cluster used in this experiment is not the most effective in big data analysis and processing due to the limitation of the number of nodes, but it shows an excellent performance improvement effect in big data processing capacity compared to ordinary single-node hosts, as shown in Figure 10. The blue line represents the average time consumption of using a single-node algorithm to process different data sizes, while the red line is the average time consumption of distributed algorithm processing. As shown in Figure 10, when the data scale is small, the traditional algorithm is even faster than the DNN algorithm, For example, in the calculation time of the first 10^4 data, the traditional algorithm is lower than the DNN algorithm, while in the $10^{th}5$ data, the DNN algorithm is still relatively stable, and the time control is 4.92 s, while the traditional algorithm exceeds 6.46 s, and the operation speed is significantly slower, and this result occurs because a certain amount of time is inevitably consumed when the Hadoop cluster starts the Map and Reduce tasks. However, as the scale of data grows, single-node algorithms begin to be unable to complete the analysis and processing of big data within effective time constraints, and the advantages of distributed
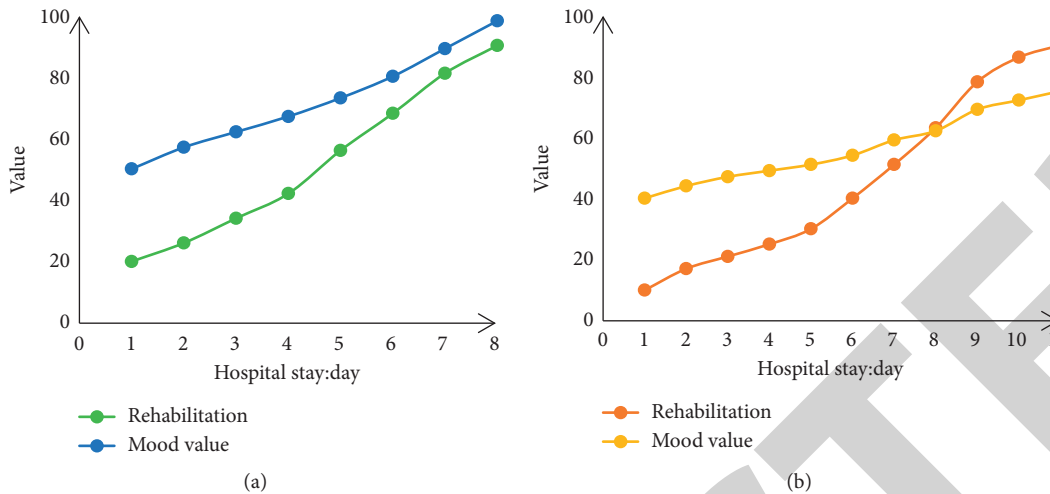
Figure 8: Postoperative rehabilitation of patients: (a) postoperative rehabilitation of patients in the test group and (b) postoperative rehabilitation of patients in the control group.

Table 4: Detailed nursing satisfaction survey.

Group	Number	Nursing errors	Nursing deficiencies	Nursing complaint rate	Nursing quality score
Control group	58	8	3	5.17	76
Test group 1	58	1	0	0	95
Test group 2	58	3	1	1.72	87
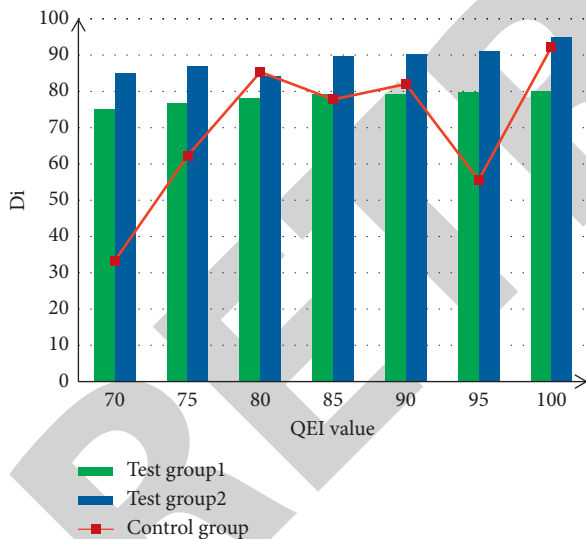
Figure 9: The relationship between quality evaluation indicators and the rate of disease conversion.
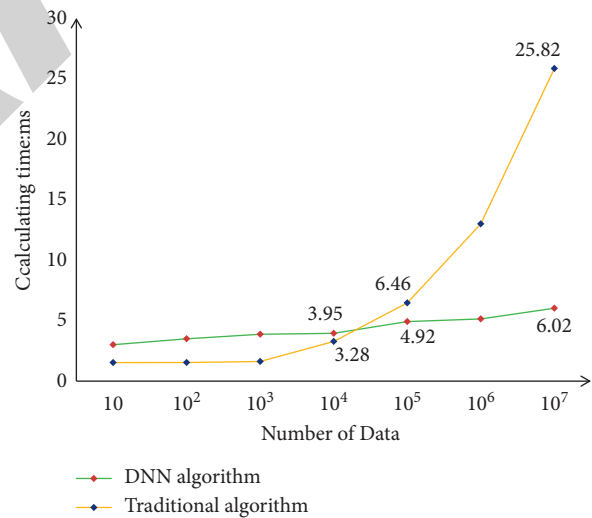
Figure 10: Comparison of DNN algorithm and traditional algorithm.

algorithms in big data processing have become apparent. In addition, when the distributed system reaches the bottleneck of data processing due to the increase in data scale, its capacity can be improved by increasing the number of DataNode nodes, which has good scalability compared with single-node algorithms.

4. Discussion

The secondary use of medical data is of great significance to medical services and clinical research, but it has not been effectively carried out due to the lack of tools and methods. A lot of knowledge and value are hidden in medical big data. Due to the lack of adequate methods, previous data analysis

and research can only rely on questionnaire surveys in medical services and clinical research, focusing on a random sampling of small-scale data such as individuals or small groups. Compared with big data analysis, these studies are not sufficiently convincing, and the methods used in these studies cannot handle big data.

The detailed nursing method and the DNN big data algorithm proposed in this article can effectively guarantee the surgical success rate of gastrointestinal tumor patients and the effect of postoperative treatment and rehabilitation in the detailed nursing and postoperative nursing in the operating room. However, due to the author's limited level, the factors considered may not be comprehensive enough. It is necessary to conduct comprehensive research from disciplines such as psychology and behavioral science to further explore the many reasons for the anxiety, fear, and other emotions of patients with gastrointestinal tumor surgery [29].

5. Conclusions

This article first analyzes and explains the background significance of detailed nursing based on medical big data and then cited the related explorations of many scholars on the subject content of this article, summarized the deficiencies of previous studies, explained the improvements made in this article, summarized the innovations studied in this article, and then deeply studied the sources of medical big data, features, application scenarios, etc. At the same time, this article proposes an improved DNN algorithm, conducts theoretical research on the algorithm, uses mathematical formulas for method calculations, and then studies the nursing process, importance, and nursing measures of detailed nursing in the operating room. Then, the design of the medical big data model is carried out. First, the relevant information of many hospitals is collected, and the relationship between the relevant parameters is analyzed by chart comparison. Then two groups of patients were selected from different hospitals for detailed nursing experiments. The final experimental results show that patients who have undergone detailed care have a 27.2% improvement in treatment and rehabilitation effects than those who have not. Moreover, the level of detailed care has an obvious positive relationship with the rate of condition conversion. In the end, the hospital's detailed care quality evaluation index, which is QEI, increases by 1 point, which can increase the rate of condition conversion by 0.4.

Data Availability

No data were used to support this study.

Conflicts of Interest

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

References

[1] A. Navarrete, D. Momblán, R. Almenara, and A. Lacy, "Giant gastric gastrointestinal stromal tumor (GIST)," *Journal of Gastrointestinal Surgery*, vol. 21, no. 1, pp. 202–204, 2017.

[2] K. K. Keditsu, S. Patkar, M. Bal, S. V. Shrikhande, and M. Goel, "Gastrointestinal neuroectodermal tumor: a diagnostic dilemma," *Indian Journal of Surgery*, vol. 79, no. 2, pp. 166–168, 2017.

[3] S. Setua, S. Khan, M. M. Yallapu et al., "Restitution of tumor suppressor MicroRNA-145 using magnetic nanoformulation for pancreatic cancer therapy," *Journal of Gastrointestinal Surgery*, vol. 21, no. 1, pp. 94–105, 2017.

[4] C. Castro Ruiz, G. Carlinfante, M. Zizzo et al., "Glomus tumor of the stomach: GI image," *Journal of Gastrointestinal Surgery*, vol. 21, no. 6, pp. 1099–1101, 2017.

[5] J.-T. Hsu, P.-H. Le, C.-F. Kuo et al., "Imatinib dose escalation versus sunitinib as a second-line treatment against advanced gastrointestinal stromal tumors: a nationwide population-based cohort study," *Oncotarget*, vol. 8, no. 41, pp. 71128–71137, 2017.

[6] S. Kenno, M. Inagaki, T. Funakoshi et al., "A case of rectal gastrointestinal stromal tumor(GIST)treated with imatinib mesylate neoadjuvant therapy to preserve the anus," *Cancer & chemotherapy*, vol. 45, no. 6, pp. 981–984, 2018.

[7] S. Nagano, N. Miyoshi, T. Takahashi et al., "Preoperative imatinib and laparoscopic intersphincteric resection for large rectal gastrointestinal stromal tumor: a case report," *International Journal of Surgery Case Reports*, vol. 71, pp. 235–239, 2020.

[8] Z. Liu, Y. Sun, Y. Li et al., "Colonic gastrointestinal stromal tumor: a population-based analysis of incidence and survival," *Gastroenterology Research and Practice*, vol. 2019, no. 10, pp. 1–10, 2019.

[9] J. P. He and J. X. Feng, "CD117 is not always positive in infantile gastrointestinal stromal tumor," *World journal of pediatrics: WJP*, vol. 14, no. 001, pp. 100–103, 2018.

[10] Y. Fukuda, T. Asaoka, H. Eguchi et al., "A case of repeated surgical resections for tumor seeding of hepatocellular carcinoma after radiofrequency ablation," *Cancer & chemotherapy*, vol. 45, no. 2, pp. 342–344, 2018.

[11] S. Hao, S. Chen, C. Tu, and T. Huang, "Anterior approach to improve the prognosis in HCC patients via decreasing dissemination of EpCAM(+) circulating tumor cells." *Journal of Gastrointestinal Surgery*, vol. 21, no. 7, pp. 1–9, 2017.

[12] M. Herzberg, M. Beer, S. Anupindi, K. Vollert, and T. Kröncke, "Imaging pediatric gastrointestinal stromal tumor (GIST)," *Journal of Pediatric Surgery*, vol. 53, no. 9, pp. 1862–1870, 2018.

[13] H. Aoyagi, H. Ito, K. Higuchi et al., "A case of fatal interstitial pneumonia during FOLFIRI plus cetuximab therapy for liver metastasis of colon cancer," *Gan to Kagaku Ryoho Cancer & Chemotherapy*, vol. 45, no. 1, pp. 51–53, 2018.

Retraction

Retracted: Analyzing the Patient Behavior for Improving the Medical Treatment Using Smart Healthcare and IoT-Based Deep Belief Network

Journal of Healthcare Engineering

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

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

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

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

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

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

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

References

[1] R. M. K. Mohamed, O. R. Shahin, N. O. Hamed, H. Y. Zahran, and M. H. Abdellattif, "Analyzing the Patient Behavior for Improving the Medical Treatment Using Smart Healthcare and IoT-Based Deep Belief Network," *Journal of Healthcare Engineering*, vol. 2022, Article ID 6389069, 8 pages, 2022.

Research Article

Analyzing the Patient Behavior for Improving the Medical Treatment Using Smart Healthcare and IoT-Based Deep Belief Network

Rasha M. K. Mohamed,[1] **Osama R. Shahin,**[2] **Nadir O. Hamed** ⓘ**,**[3] **Heba Y. Zahran** ⓘ **,**[4,5,6] **and Magda H. Abdellattif** ⓘ[7]

[1]*Department of Chemistry, College of Science, Jouf University, P.O. Box 2014, Sakaka, Saudi Arabia*
[2]*Physics and Mathematics Department, Faculty of Engineering, Mataria, Helwan University, Egypt*
[3]*Computer Studies Department, Elgraif Sharg Technological College, Sudan Technological University, Khartoum, Sudan*
[4]*Laboratory of Nano-Smart Materials for Science and Technology (LNSMST), Department of Physics, Faculty of Science, King Khalid University, P.O. Box 9004, Abha 61413, Saudi Arabia*
[5]*Research Center for Advanced Materials Science (RCAMS), King Khalid University, P.O. Box 9004, Abha 61413, Saudi Arabia*
[6]*Nanoscience Laboratory for Environmental and Biomedical Applications (NLEBA), Metallurgical Lab. 1, Department of Physics, Faculty of Education, Ain Shams University, Roxy, Cairo 11757, Egypt*
[7]*Department of Chemistry, College of Science, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia*

Correspondence should be addressed to Nadir O. Hamed; nohamedit@gmail.com

Patient behavioral analysis is a critical component in treating patients with a variety of issues, with head trauma, neurological disease, and mental illness. The analysis of the patient's behavior aids in establishing the disease's core cause. Patient behavioral analysis has a number of contests that are much more problematic in traditional healthcare. With the advancement of smart healthcare, patient behavior may be simply analyzed. A new generation of information technologies, particularly the Internet of Things (IoT), is being utilized to transform the traditional healthcare system in a variety of ways. The Internet of Things (IoT) in healthcare is a crucial role in offering improved medical facilities to people as well as assisting doctors and hospitals. The proposed system comprises of a variety of medical equipment, such as mobile-based apps and sensors, which is useful in collecting and monitoring the medical information and health data of patient and interact to the doctor via network connected devices. This research may provide key information on the impact of smart healthcare and the Internet of Things in patient beavior and treatment. Patient data are exchanged via the Internet, where it is viewed and analyzed using machine learning algorithms. The deep belief neural network evaluates the patient's particulars from health data in order to determine the patient's exact health state. The developed system proved the average error rate of about 0.04 and ensured accuracy about 99% in analyzing the patient behavior.

1. Introduction

Health provides an improvement in quality of life by using smart healthcare application like diagnosis and treatment of disease, smart pills, remote patient monitoring, biosensors, emergency healthcare, robotics, and healthcare facilities in 24/7/365 [1–6]. In digital world, everything is becoming informative and getting transformed. Traditional biotechnology and medicine may now become digital and become more informed thanks to technological advancements and increasing expansion. The combination of smart healthcare has resulted in the emergence of new information technology. The smart health-care system is a multilevel transformation. The patient behavior is analyzed by making

changes in smart healthcare system including a shift from disease-centered to patient-centered treatment, a shift in management and prevention concepts, and a shift in information creation from regional to clinical informatization [2, 7, 8].

For achieving the aforementioned aims, current advancements in Internet of things (IoT) technology allow scalable and less expensive linked sensor networks. Several review papers on IoT technology, architecture, and applications in healthcare have been published. Massive volumes of data are accessible from the healthcare IoT, which may be used for descriptive, diagnostic, predictive, and prescriptive analytics [9, 10].

The Internet of Things (IoT) is the interconnection of devices, network connections, and sensors which permits these things to collect and share information more efficiently. The continual monitoring of a patient through numerous metrics and inferring a favorable outcome from the history of such constant monitoring is the defining feature of IoT in the healthcare system. In today's ICUs, several of these gadgets with medical sensors are present [5]. Despite 24 hours continuous monitoring, there may be times when the doctor is not informed in time when there is an emergency. There may also be difficulties in exchanging data and information with specialists as well as concerned family members and relatives. The technology to improve these features is currently available, but it is out of reach and out of reach for the majority of people in developing nations like India. As a result, these remedies to these difficulties might be as easy as adding these features to current devices that do not have them.

Nowadays, IoT is the effective technologies which is rising and allows computing device and things to communicate with one another and transport information from one location to another. More importantly, the IoT sensor aids in communication improvement without necessitating human-computer or human-human contact. Hospital systems, smart home maintenance, covering devices, thermostats, home appliances, lighting fixtures, and security systems have all benefited from the use of IoT sensor devices [6]. More ecosystems, including smartphones and smart speakers and smartphones, were successfully coupled with these applications. The IoT-based sensors played a critical role in the health business, which has a high number of healthcare services, among the many applications. The improvement of technology and the growth of worldwide populations resulted in an upsurge in the number of chronic illnesses, which were often problematic to detect and were also more expensive. The most common healthcare service was sometimes out of reach for a large number of individuals and communities, resulting in a rise in the chronic illness impacted ratio. The majority of respondents believed that illness diagnostics and patient assistance systems required more resources. They also believed that healthcare services were not packet friendly, implying that they were inaccessible. As a result, IoT-based services have been supplied to the healthcare industry in order to improve people's generosity in terms of availability.

The IoT utilization in healthcare industry has increased common hope between patients and service providers, as well as the effectiveness of treatment plans [7]. Furthermore, the IoT-based illness diagnostic procedure significantly improved the quality of therapy. Simultaneous reporting and monitoring, cost, end-to-end connection, remote medical aid, data analysis, tracking, and alert were all advantages of IoT sensor-based healthcare services.

The workflow of patient was effectively sustained by employing active technologies and facilities, which was the key cause for integrating IoT devices in health equipment. Data transfer, data interoperability, information interchange, and machine-machine communication all improved as a result of the technologies used. Patients' needless visits were decreased as a result of better use of technology and communication protocols, which improved planning and allocations [8]. Furthermore, the gadget itself had a data-driven analysis tool that formed a graph based on variances, allowing for more accurate decision-making. When the collected medical data revealed any potentially dangerous actions, the medical gadget sent out an on-time notice. The alarm-related information was sent to the doctor in order to create a real-time tracking system and deliver attentive messages to the connected system [9]. The appropriate therapy was given in response to the alarm, and the choice was made effectively to extend people's lives. According to the debate, the IoT sensor gadget successfully observed the patient's behaviors, notifying, tracking, offering better therapy, and entirely improving patient care. It stated unequivocally that the IoT sensor gadget was solely utilized to monitor the patient in a distant activity. The remote-based action assured that the patient's therapy was effective, as well as providing medications depending on the disease.

2. Related Work

The modified particle swarm optimization technique was used to create an efficient IoT device-based healthcare system. Using the previously indicated optimization strategy in healthcare applications, the remote identification system was established throughout this procedure. Inertia weight factors, shrinkage factors, and other data fusion characteristics were all utilized by the system. The performance of particle swarm optimization techniques was optimized using these parameters. This enhanced system correctly forecasted patient health data from numerous sites, allowing for a better understanding of a given person's behavior [10].

A mobile health app was developed that uses IoT sensors to forecast sickness in the healthcare industry. During this procedure, the UCI repository dataset was acquired via a mobile application. The fuzzy rule-based neural classifier was used to process the acquired data, and it was successful in detecting the sickness that was present in the human body. When compared with other traditional approaches, the created system accurately anticipated the changes in the patient with the best accuracy of 98.5% [11].

It is important to recognize that behavior analytics is a broad term that encompasses descriptive, diagnostic, predictive, and prescriptive analytics. Descriptive analytics

might be utilized to examine several healthcare decisions and their possessions on application performance and outcomes. The information is often offered in simple tables, graphs, and charts that show important indicators like heart rate and temperature as a purpose of time. Medical diagnosis is a term that may be used to describe diagnostic analysis [12].

A hybrid random forest with a linear model used as an efficient IoT-based heart disease diagnostic system is developed. Smart devices were used to capture cardiac data, which was then analyzed using optimal machine learning algorithms to correctly forecast cardiovascular illness. The use of an IoT device-based examination method decreased the challenges of the clinical examination process and aided in making an informed conclusion. When it came to identifying cardiovascular illness using IoT sensor data, the developed system had an accuracy of 88.7% [13].

Deep neural network techniques were used to construct a real-time health monitoring system based on IoT sensor. The system was completely focused on analyzing real-time ECG data in order to forecast variation in heart activity and determine cardiovascular state. Furthermore, the technology was capable of increasing illness identification rates at a cheaper cost [14]. The successful creation in the ECG monitoring system based on IoT decreases the complexity of the treatment process and enhances the connection among the healthcare sectors and patient.

2.1. Smart Technology.

The information and communication technology are used to refer the smart technology in healthcare, as well as health monitoring system, and it encompasses a wide range of technologies, such as wearable, and smart house technology. A computer, mobile phone, voice activation system, touchpad controller, Personal Digital Assistant, and other device such as a remote control are all examples of smart technology that can interface with and operate the devices in a person's working environment. The system may be controlled by infrared extension units, radio frequency, and sound [15].

Smart phones, which have various functions beyond phone calls, are one of the most widely utilized smart technologies. As a result, it is critical to explore the utilization of smart phones for communications, such as calling a medical practitioner, and for assisting persons with their medical requirements, such as treatments for sleep management concerns or glucose level maintenance in diabetics.

For many years, technology has been employed in healthcare to increase testing, imaging, and surgical expertise. There are more advanced services and healthcare delivery innovations used to provide patient care in the fast-developing technological environment of medicine. Furthermore, advances in vital medical equipment have expanded physicians' and patients' diagnostic and treatment choices. Medical practitioners have enhanced the quality of care in the areas of screening and prevention, as well as diagnosis and treatment, by using technology [16]. Extrapolation of these achievements to patients and caregivers at home and in the community is an ongoing and critical

component of medical intervention. As healthcare professionals, insurers, and patients see the benefits and efficiency of these integrated models of care, the usage of telehealth technology and smart devices in day-to-day patient care is rapidly increasing.

2.2. IoT-Enabled Healthcare.

The positioning of the healthcare industry, which comprises IoT services and IoT apps, as well as solutions, drives highly innovative connected health technolines. The fundamental goal of digital health is to improve healthcare services while lowering costs. All distant patients can receive diagnosis and treatment via wearable devices in smart healthcare [17]. The patient's condition is assessed using gadgets which are incorporated with the IoT-based sensor. Wearable devices gather the patient's pulse rate, oxygen saturation, and blood glucose and send them to the caregiver via the patient's smartphone. Doctors can consult with patients via telehealth instead of visiting to a clinic or hospital. The behavior change is a practice which may help people improve their health by altering unhealthy behavior and adopting healthier lives.

2.3. Importance of Smart Healthcare System.

Hospitals, patients, doctors, and research organisations are only a handful of the healthcare system's players. Smart healthcare is built on the foundations of mobile Internet, IoT, 5G, big data, cloud computing, artificial intelligence, and biotechnology. These technologies are employed in the healthcare system to provide smart healthcare. Patients who use wearable gadgets use them to keep track of their health at all times, offer medical treatment through virtual assistants, and perform remote services from distance; doctors use a range of sophisticated clinical judgments to support systems to enhance diagnosis [18]. Doctor's forms handle integrated information platforms, which include tools like communication systems, photo archiving, laboratory information management systems, and the electronic medical record. The surgery robots are useful to assist the surgery with mixed reality technologies.

Radiofrequency identification (RFID) is used to track people and their belongings, as well as the hospital's supply chain. RFID is used in conjunction with combined management systems to gather data and aid decision-making. The utilization of mobile medical platforms to improve mobile medical platforms is common [19]. Machine learning is frequently employed in scientific research technology instead of manual drug screening, and big data may be used to find appropriate individuals. The smart healthcare technology helps in expense, and danger of medical operations may be substantially minimized. By integrating IoT into the smart healthcare system, tailored medical services such as resource utilization efficiency, self-service medical care, and regional exchange and collaboration may be provided.

3. Materials and Methods

The materials and methods utilized in this paper for examining the behavior of patients using smart healthcare are based on technological advancements. A detailed survey on

IoT-based healthcare has been conducted. IoT-enabled healthcare entails the creation of an IoT healthcare network, which is a crucial component of a smart healthcare system. In IoT healthcare network platform, IoT healthcare network architecture, and IoT healthcare network topology make up an IoT healthcare network. IoT-enabled healthcare is separated into several services and applications, with each service having its own potential sector [20]. There were about 400 clinical leaders and healthcare professionals in attendance.

The smart healthcare technology described in this paper is proficient of making decisions based on the detected circumstances of a patient's pulse rate, heartbeat, and body temperature. The sensors, as well as the system's cost and longevity, are managed by the architecture's algorithm. The difficulties encountered in remote patient monitoring aid in the provision of appropriate therapy by professional doctors in the hospital. The smart healthcare monitoring system and patient management system are monitored through communication channels and incorporated internal and exterior sensors [21]. Telemedicine improves the operating performance of telemedicine in rural regions. The Internet of Things has aided in the implementation of the gadget at a rural clinic. The gadget collects all patient data and sends it to the appropriate hospital doctor. These data are analyzed by doctors, who then recommend certain actions for the patient's optimum therapy.

The smart healthcare system makes use of a variety of devices. The smart system for healthcare was developed using a room temperature sensor, body temperature sensor, an ESP32 processor, a heartbeat sensor, a CO sensor, and a CO2 sensor. The potentiometer may be used to modify the sensor's sensitivity. The sensors used in air quality management detect and quantify NH3, smoking, nicotine, benzene, and CO2. The sensor module features a digital pin that enables it to function without a microcontroller, which is beneficial for detection of various gases [22]. Any organ that controls the amount of light that passes through it may detect changes in blood volume very fast. The pulse timings are far more significant in devices that assess heart rate. When light is expended by the blood, the heartbeat rate has an influence on signal pulse and blood volume distribution. This mechanism is similar to heartbeat pulses. The proposed system is illustrated in Figure 1.

A remote health monitoring system is controlled by Raspberry Pi. Raspberry Pi is a powerful and small-sized microprocessor designed to improve the properties of computer science teaching in colleges as well as the remote health monitoring systems. The aim of this study is to progress a system that can continually monitor vital indicators such heart rate, pulse rate, blood pressure, and body temperature. The data are saved on a cloud server database and can only be accessed by authorised people via an Internet website or mobile application. Although the concept is not new, it is offered as an absolute and inexpensive approach for the system utilising the Raspberry Pi. The major goal of this system is to update data online and provide alerts to clinicians if anything is odd, as well as to anticipate if the patient has a condition.
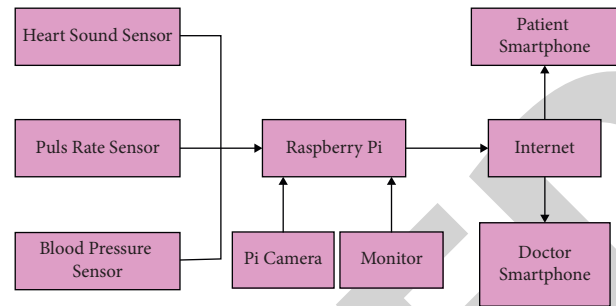

Figure 1: Block diagram of the IoT-based smart technology.

3.1. IoT-Based Deep Belief Neural Network.

Deep belief neural networks may be taught using a variety of learning algorithms. To ease the process, a neural network learning technique is employed to choose training data. The output data of a deep belief neural network are handled through layers and weights to reduce the uneven network output to layers. In the graphic below, a deep belief neural network for data processing in the IoT industry is depicted [23]. The feed-forward technique is used in the deep belief neural network, and the layers are coupled to each node in a multilayer perceptron. The handling of IoT sensor data is a hazy process involving a large number of data on the server side. In order to make decisions, the deep belief neural network's learning process comprises training data samples. Figure 2 examines the processing unit of IoT.

The weights of the neural network capture the variables of the neural network, transforming the data to decision-makers. To observe the expected error in the network, repeated training data are utilized. The input signal is sent to the input network's equivalent network, which creates an output containing errors and noisy data. Because the neural network seems to have distinct parameters, partial derivatives of each variable are employed to lower the error rate. The value is then sent via the activation function, where specific weights multiply the amounts from the hidden layer to the output layer, and the output sums up the input it is getting before sending the result over the activation function and produces output. The output layer yield is then compared to the target output in a neural network [22–25].

The deep belief neural network has the ability to train the feature of patient using the layer of unsupervised learning technique to improve the monitoring process of patient with IoT sensor which is shown in Figure 3. In training process, while applying learning weights, the maximum likelihood value is used in the model. The weighted value is updated using the gradient descent function. The sensor in IoT device is processed to every neuron in the deep belief neural network, in which the data aggregation and processing takes place.

The input signal x_i is shared in which the neuron layer of input serves as the buffer. The output y_i is evaluated form the input layer with the weighted connection $\varnothing_{ij}$. The input signal is added to the intermediated layer and strengthens the weighted connection. The output layer of the neural network is defined by
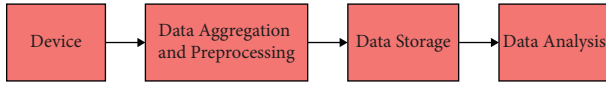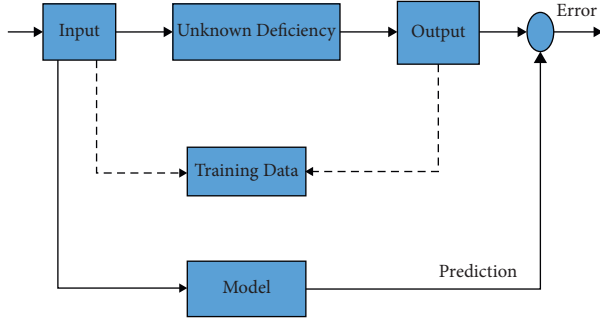
FIGURE 2: IoT-based data processing unit.

FIGURE 3: Deep belief neural network for IoT smart health monitoring.

$$y_j = f(\text{net}_j),$$
$$\text{net}_j = \sum_i \varphi_{ij} x_i + \varnothing_i, \quad (1)$$

where $\varnothing_i$ is the bias variable. The range of output y_j is from −0.1 to 0.1 which defines the frequent function regard the input. The output data in deep belief neural network are managed to minimize the output layer of the network. The error E is expressed as

$$E = \sum_s E_s,$$
$$E = \frac{1}{2} \sum_s \sum_i (dx_{si} - x_{si})^2. \quad (2)$$

The space parameter is determined by taking partial derivates of E:

$$\frac{\partial E}{\partial \varnothing_i} = \sum_s \delta_{si},$$
$$\delta_{si} = -\frac{\partial E_s}{\partial x_{si}} \frac{\partial x_{si}}{\partial \text{net}_{si}}. \quad (3)$$

Then,

$$\frac{\partial E_s}{\partial x_{si}} = - \sum_k \delta_{sk} \varnothing_{ki}. \quad (4)$$

This demonstrates the output layer of the processed IoT data.

Procedure: Deep belief neural network using supervised learning (x_i, y_j)
f = (x_i, y_j)
Initialize samples and data to train
Repeat
For all $(x_i, y_j)^E$ do
Compute E_s
Compute δ_{si}
Compute $\partial E_s / \partial x_{si}$
Update the value of $\varnothing_i$
End for
Repeat until achieving all samples
End

4. Result and Discussion

A smart healthcare system enables doctors to study a patient's behavior, which aids in the provision of an accurate diagnosis. Patients may check their health condition at any time and at any location thanks to the smart health management system's monitoring. Smart healthcare systems make this possible by putting this information on a digital platform that can be accessed from anywhere in the globe. The smart healthcare system allows for the real-time value of the health system and also demonstrates how it may be applied in the actual world. The smart healthcare system can readily identify the patients' diseases, and diagnosis may be done by watching the numerous symptoms. Doctors can utilize the record of the patient's physical states to analyze the medicine's influence on other things thanks to the smart healthcare system. The approach was put to the test with a diverse group of contributors of various ages in a range of settings. In the test situations for heartbeat, a room temperature, and body temperature sensor, the true value and observed value were manually determined from the proposed system. The room temperature sensor is solely utilized to calculate humidity; in this example, error rate has been determined using the data to indicate the system's efficiency. The deviation in room humidity is examined in Figure 4.

Th error rate is calculated based on the system effectiveness. The error rate of body temperature, heart rate, and room humidity is given in Figure 5 and Table 1. The error is due to the deviation in the patient treatment, which leads to the inaccurate data.

Table 2 and Figure 6 are given based on the data collected based on the room humidity. The smart healthcare IoT system observes the data which may vary sometimes due to the error in the IoT sensor system.

Table 3 provides the data collected form the patient to analyze the behavior and providing the treatment using the IoT smart health monitoring system illustrated in Figure 7. The actual value and the observed value may vary because of the error and deviation in sensor.

The patient behavior detection provides the efficient performance using weight selection process and retraining process with less error rate. The error rate is graphically depicted in Figure 8. The effective pretraining, unit vector for weight updating, and hidden unit is useful for selecting the efficient feature for monitoring the patient activities. The deviation of estimated value and computed value is minimum in deep belief network because of effective pretraining.

The efficiency in the selection process is estimated using the precision value which is given as follows:
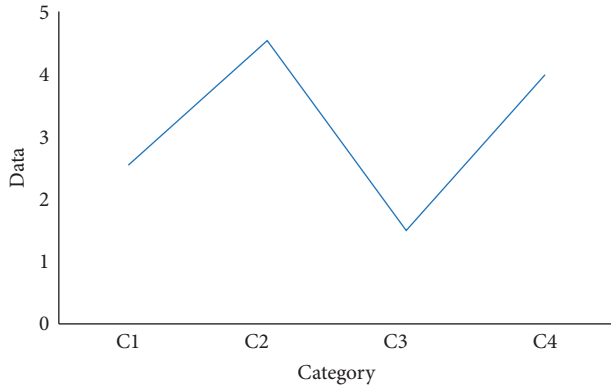
Figure 4: Deviation in room humidity.

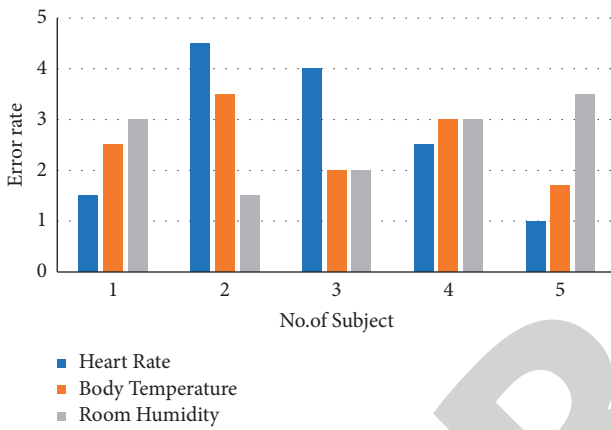Figure 6: Data collected for room humidity.

Table 3: IoT-based smart health monitoring system.

Subject	Actual data	Observed data
1	64	66
2	68	71
3	74	75
4	73	73
5	71	70

Figure 5: Error rate of the developed system.
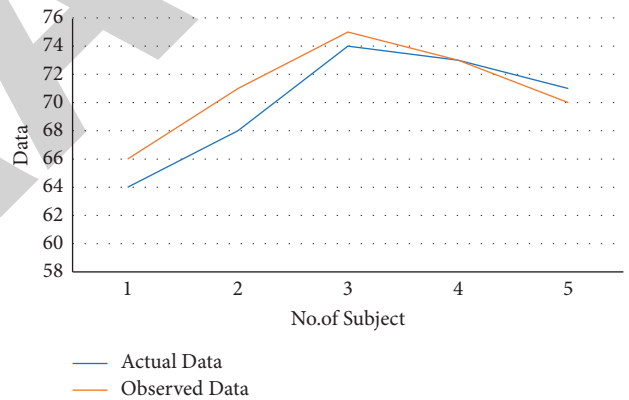
Figure 7: Data observed using the IoT smart health monitoring system.

Table 1: Observed error rate.

Subject	Error rate		
	Heart rate	Body temperature	Room humidity
1	1.5	2.5	3
2	4.5	3.5	1.5
3	4	2	2
4	2.5	3	3
5	1	1.7	3.5

Table 2: Data collected for room humidity.

Subject	Actual data	Observed data
1	63	61
2	66	67
3	63	62
4	68	69
5	64	62

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{5}$$

Deep belief network achieves high value compared with other classification techniques. The patient behaviors detection provides the effective performance using the weight selection and retraining process with maximum selection value. The precision value is depicted in Figure 9. The value

of the monitoring patient is obtained from the training input data. The deep belief network has high precision value and provides effective performance on iterative weigh selection and enhances the patient behavior analysis with maximum precision value.

The effectiveness of the smart monitoring system based on IoT is estimated by utilising the accuracy metrics. The accuracy value is estimated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{6}$$

Figure 10 illustrates the accuracy value of deep belief neural network. This network provides high accuracy compared with the existing techniques. Pretraining, weight update, learning, and activation functions all help to enhance the overall performance of the patient monitoring system based on the IoT sensor.

FIGURE 8: Error rate prediction using deep belief neural network.

FIGURE 9: Validation of Precision rate.

FIGURE 10: Accuracy rate.

5. Conclusion

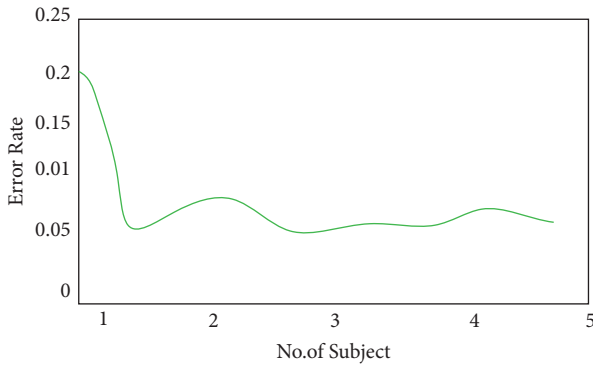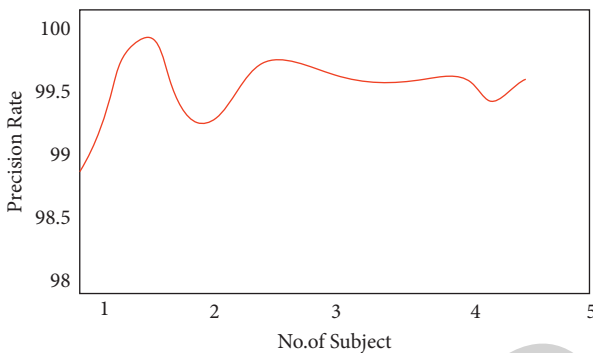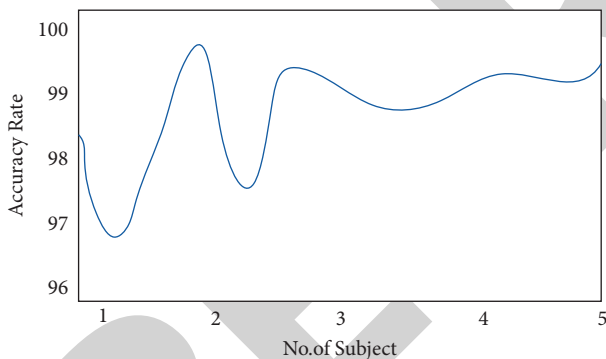This paper concludes that the Internet of Things has the potential to play a critical role in the advancement of smart healthcare. The Internet of Things improves the flexibility of the healthcare system. It is critical to incorporate Internet of Things in order to deliver exact diagnosis to patients. The proper architecture is present in the smart healthcare system's growth and development. It has been determined that smart healthcare plays a significant role in analyzing patient behavior. Patients can benefit from smart healthcare for better health self-management. Smart healthcare systems as related to traditional healthcare can minimize staff pressure, lower expenses, and enhance the patient's medical practice. The smart healthcare can help to improve the status of

medical resources. This research may provide key information on the impact of smart healthcare and the Internet of Things in patient behavior and treatment. Patient data are exchanged via the Internet, where it is viewed and analyzed using machine learning algorithms. The deep belief neural network evaluates the patient's particulars from health data in order to determine the patient's exact health state. The developed system proved the average error rate of about 0.04 and ensured accuracy about 99% in analyzing the patient behavior.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

References

[1] V. Jagadeeswari, V. Subramaniyaswamy, R. Logesh, and V. Vijayakumar, "A study on medical Internet of Things and Big Data in personalized healthcare system," *Health Information Science and Systems*, vol. 6, no. 1, pp. 14–20, 2018.

[2] C. Rathert, E. S. Williams, D. McCaughey, and G. Ishqaidef, "Patient perceptions of patient-centred care: empirical test of a theoretical model," *Health Expectations*, vol. 18, no. 2, pp. 199–209, 2015.

[3] A. Ismail, M. L. Salem, A. Elkholy, W. Elmashad, and G. A. M. Ali, "In-silico analysis of protein receptors contributing to SARS- COV-2 high infectivity," *Information Sciences Letters*, vol. 10, no. 3, pp. 561–570, 2021.

[4] O. Y. M. Al-Rawi, W. S. Al-Dayyeni, and I. Reda, "COVID-19 impact on education and work in the kingdom of Bahrain: survey study," *Information Sciences Letters*, vol. 10, no. 3, pp. 427–433, 2021.

[5] E. G. M. Petrakis, S. Sotiriadis, T. Soultanopoulos, P. T. Renta, R. Buyya, and N. Bessis, "Internet of things as a service (itaas): challenges and solutions for management of sensor data on the cloud and the fog," *Internet of Things*, vol. 3-4, pp. 156–174, 2018.

[6] M. Alrizq, S. A. Solangi, A. Alghamdi, M. A. Nizamani, M. A. Memon, and M. Hamdi, "An architecture supporting intelligent mobile healthcare using HumanComputer interaction HCI principles," *Computer Systems Science and Engineering*, vol. 40, p. 558, 2022.

Research Article

Machine Learning Technology-Based Heart Disease Detection Models

Umarani Nagavelli,[1] **Debabrata Samanta** (ID),[1,2] **and Partha Chakraborty** (ID)[3]

[1]*Dayananda Sagar Research Foundation, University of Mysore (UoM), Mysore, Karnataka, India*
[2]*Department of Computer Science, CHRIST (Deemed to be) University, Bangalore, Karnataka, India*
[3]*Department of Computer Science and Engineering, Comilla University, Cumilla-3506, Bangladesh*

Correspondence should be addressed to Partha Chakraborty; partha.chak@cou.ac.bd

At present, a multifaceted clinical disease known as heart failure disease can affect a greater number of people in the world. In the early stages, to evaluate and diagnose the disease of heart failure, cardiac centers and hospitals are heavily based on ECG. The ECG can be considered as a regular tool. Heart disease early detection is a critical concern in healthcare services (HCS). This paper presents the different machine learning technologies based on heart disease detection brief analysis. Firstly, Naïve Bayes with a weighted approach is used for predicting heart disease. The second one, according to the features of frequency domain, time domain, and information theory, is automatic and analyze ischemic heart disease localization/detection. Two classifiers such as support vector machine (SVM) with XGBoost with the best performance are selected for the classification in this method. The third one is the heart failure automatic identification method by using an improved SVM based on the duality optimization scheme also analyzed. Finally, for a clinical decision support system (CDSS), an effective heart disease prediction model (HDPM) is used, which includes density-based spatial clustering of applications with noise (DBSCAN) for outlier detection and elimination, a hybrid synthetic minority over-sampling technique-edited nearest neighbor (SMOTE-ENN) for balancing the training data distribution, and XGBoost for heart disease prediction. Machine learning can be applied in the medical industry for disease diagnosis, detection, and prediction. The major purpose of this paper is to give clinicians a tool to help them diagnose heart problems early on. As a result, it will be easier to treat patients effectively and avoid serious repercussions. This study uses XGBoost to test alternative decision tree classification algorithms in the hopes of improving the accuracy of heart disease diagnosis. In terms of precision, accuracy, f1-measure, and recall as performance parameters above mentioned, four types of machine learning (ML) models are compared.

1. Introduction

Cardiovascular disease (CVD) is a type of heart disease that continues to be a major cause of death worldwide, accounting for over 30% of all deaths. If nothing is done, the total number of fatalities in the world is anticipated to rise to 22 million by 2030. Plaques on arterial walls can obstruct blood flow, resulting in a heart attack or stroke. Heart disease is caused due to various risk factors such as physical inactivity, unhealthy diet, and the effective use of alcohol and tobacco [1, 2]. The abovementioned factors are reduced by adopting a good daily lifestyle, namely, reducing salt in the diet, consumption of vegetables and fruits, practicing physical activity regularly, and discontinuing alcohol and tobacco use, which helps to minimize the risk of heart disease [3]. The solution to overcome these problems is to use the collection of patient records from different health care centers and hospitals. For getting the results and seeking another opinion from an experienced doctor the decision support system is used. The unnecessary test conductions are avoided by this technique for diagnosis, thereby saving money and time [4, 5]. Recently, a system of hospital management was utilized for managing the health care or patient data which means more data are produced

by these systems. For predicting heart disease, the DSS utilizing the NB (Naïve Bayes) algorithm was developed. A web application is constructed to obtain the application and user input, and it retrieves key features relating to heart disease from a historical database (the Cleveland data set) [6, 7].

More number of neurohormonal regulatory mechanisms are triggered in the initial stages of heart failure disease (HFD). In a short duration, these compensatory mechanisms can cause the HFD consequences, leading to accentuated ventricular dysfunction, dyspnea on exertion, peripheral edema, pulmonary, and heart remodeling which can cause afterload and preload permanent changes. More options of treatment with HFD are given to the patient including lifestyle changes and implantable or medication devices such as a defibrillator or pacemaker. The main concern is ensuring the follow-up in this population given that hospitalization due to acute HFD decompensation is the leading cause of healthcare expenditure. Statistics and studies show that heart diseases are the most significant issue faced by people particularly HFD [8, 9]. For various diseases, early diagnosis and detection of cardiac disease is the first step in care and treatment.

The HFD is now an emerging disorder for diseases such as hypertension, insomnia, and heart disease among others. The HFD detection on ECG is completed through variations detection in duration of heart beats from the time interval from 1 wave of PQRST to the next wave of PQRST. For IHD early detection, an emerging and promising noninvasive diagnostic tool is MCG (Magenetocardiography). While MCG is less influenced by contact interference of electrode-skin compared to ECG, it is highly sensitive to vortex current and tangential causes through the tissue of ischemic cardiac. Despite its high signal quality, MCG interpretation is time-consuming, highly dependent on interpreting experience, and has limited appeal in clinics. As a result, clinicians would benefit from an autonomous system that can detect and localize ischemia at an early stage [10].

Early identification of heart disease of improved diagnosis and high-risk individuals using a prediction model can be recommended generally for fatality rate reduction, and decision-making is improved for further treatment and prevention. In CDSS, a prediction model is implemented and utilized to support the clinicians in assessing the heart disease risk, and appropriate treatments are provided for managing the further risk. Additionally, numerous studies have also reported that CDSS implementation can improve decision quality, clinical decision making, and preventive care, respectively [9, 11]. Coronary artery disease (CAD), also known as ischemia heart disease (IHD), is the leading cause of death in adults over the age of 35 in different countries. During the same time span, it became China's biggest cause of death. When blood flow to the heart is reduced due to coronary artery stenosis, IHD occurs. Myocardial damage can have serious consequences including ventricular arrhythmia or even sudden cardiac death due to myocardial infarction.

1.1. Major Contribution of Research.
Machine learning may be used to diagnose, detect, and forecast many disorders in the medical industry. The primary purpose of this study is to give clinicians a tool to detect cardiac problems at an early stage. As a result, it will be easier to deliver appropriate treatment to patients while avoiding serious effects. This study uses XGBoost to test several decision tree classification algorithms in the hopes of improving performance in heart disease diagnosis.

The remaining paper is organized as follows: In Section 2, literature analysis is presented. ML techniques for heart disease prediction is presented in Section 3. Feature extraction is projected in Section 4. Heart disease prediction using the XGBoost algorithm is elaborated in Section 5. Results are shown in Section 6. The concluding remarks are given in Section 7.

2. Literature Analysis

In the system of the human heart, the heart's electrical activity is recorded by ECG with various wave forms through skin electrodes. For heart disease identification, it is a noninvasive technique which reflects heartbeat, heart rate, and cardiac health. The number of cells in the human body has no direct contact with the outer location. Moreover, they depend on the cardiovascular system for serving as a provision of transport for them. In system of cardiovascular, the fluids are two kinds are flow through it. Blood is the first fluid type. Here, the circulatory system forms the blood vessels and heart. Lymph is the second type of fluid. The structure of the lymphatic system is formed by lymph nodes and lymphatic vessels. The cardiovascular system can be formed by the combination of the lymphatic system and the vascular system [12]. In heartbeat, a heart cycle is an action series. A heart cycle typically contains both atria, with each ventricle contraction synchronized a fraction of a second later. The heart is produced and interconnected using heart muscle cells, so when one of them contracts, it causes nearby cells to excite. The muscles rest between beats in the cardiac cycle, which contributes to aerobic breathing. Two parts are there, which are further discussed in this study.

2.1. Part 1 Is 'Systole'.
It is the expression for contraction. It happens whenever the ventricles are in the stage of contracting and cause the blood to pump into the vessels of the heart with A–V valves closure and semilunar valves opening.

2.2. Part 2 Is 'Diastole'.
It is an expression for relaxation. It happens whenever the ventricles are in a stage of relaxing. This causes back-blood pressure to the valves of the closing semilunar and opening valves of A–V, respectively.

Previous studies have shown promising results for CVD automatic detection. But still, there are a few concerning problems. Firstly, studies utilizing private datasets suffer from variations of database and small-sized databases, especially for MCG studies in which larger public datasets are unavailable. However, based on public datasets, ECG studies might not accomplish equal performance if they are

transferred from benchmark to clinical domain. Representatives composes the public datasets of ECG. The ECG datasets easily identify abnormal cases and would be biased to perform early diagnosis [13]. On model performance, feature selection has shown a significant impact. Past studies presented features from three categories: (a) information theory features, (b) time domain features, and (c) frequency domain features.

Every category is represented as effective and has the ability for revealing some complicated aspects of cardiac electrophysiological signals. Certain studies are carried out to evaluate the significance of different features. For solving this issue a large feature group is relatively designed that contains abovementioned two categories. Through the feature importance weight analysis, the following is determined:

(i) Repolarization synchronicity of T wave is described by features as a core feature for identifying the subjects of IHD

(ii) Features describe that the characteristics of the magnetic pole are associated with coronary stenosis locations

Various studies are reporting that the heart disease diagnosis development based on models of ML can provide the objective of HDPM with improved performance. The 2 publicly available datasets of heart disease such as Cleveland and Statlog are mostly used for comparing the prediction models' performance between researchers. In the area of healthcare, ML-based clinical decision making has been applied recently. In machine learning, recent advances are representing discriminative classifiers' advantages for cardiac disease automatic detection. Studies have previously shown that machine learning algorithms, namely, SVM, RF (random forest), LR (logistic regression), BPNN (back propagation neural network), and MLP (multilayer perceptron) are utilized successfully for decision-making tools to predict heart disease based on individual information. Various studies revealed the hybrid model merits that achieved good performance in heart disease prediction, namely, RF with a linear model, MLP, Bayes Net (BN), majority voting of NB, and RF and two stacked SVMs, respectively [14]. Kalia Orphanou et al., in the NB classification model, the TARs (Temporal Association Rules) feature is used for diagnosing heart disease. To preprocess the data, a temporal abstraction (TA) is used and a temporal pattern mining algorithm is utilized for finding TARs by frequent temporal relationships identification between TAs. In Naïve Bayes classifier, periodic TARs are considered as features finally. With the help of possible recurrence of each TAR pattern with relevant patient history, an 82% accuracy is obtained. Theresa Princy et al. have conducted a survey on several machine learning techniques which are utilized to predict the heart diseases risk of a person depending on various attributes such as gender, age, cholesterol, and pulse rate. When increasing the accuracy of risk and attributes, the author can use the analysis of the K-nearest neighbour algorithm, Naïve Bayes, and neural network. The accuracy is increased with a low number of attributes, which is possible by using various methods [15]. Prerana et al. predicted the risk level of probabilistic analysis and classification (PAC) and heart disease completed by machine learning technique. For handling patients' records and various machine learning algorithms analysis, the map reduce programming is used and given with the graphical representation. In the cloud, this approach is available and globally accessible. Furthermore, it can be extended for determining different diseases such as cancer, diabetes, brain tumor, and so on. Shadab et al. use the technique of NB data mining for supporting the users to know the answers for predefined questions in the application of web based. For diagnosing heart disease, doctors use these intelligent decisions when NB algorithm accuracy can be improved by utilizing various techniques [16]. Figure 1 shows ML techniques for heart disease prediction.

3. ML Techniques for Heart Disease Prediction

3.1. Naïve Bayes Weighted Approach (NBwa).
The web application is the proposed system. It can be classified into two modules: admin side and user side, respectively. The admin submitted the dataset into the database in the form of an excel sheet, which was then translated into weights and stored. Clusters were formed from the records. Depending on average weight or each of the clusters is determined, attribute (vessel) classification and in the database this feature can be stored. Using provided fields, the user inputs and patient data are converted into weights, and the algorithm of NB is applied to the data. The result can be displayed to the user based on system probability. Various events' conditional and marginal probabilities are compared by the Naïve Bayes algorithm. For the given samples, this algorithm is useful for calculating the possible nearest value [17]. The Bayes theorem is useful for calculating the diagnostic probability when the patient's health is monitored based on a few symptoms. The Bayes rule is used in a variety of data mining approaches. This technique is more beneficial for developing predictive capability models and provides a variety of ways to examine and explore the data. The Naïve Bayes classifier is an appropriate scheme when a given input attribute is more. Compared to many other classification techniques this algorithm is simple, but it has better performance. The heart disease patient's characteristics are easily identified by the Naïve Bayes classifier technique. This algorithm will find the input attributes' probability during a predictable disease state [18]. Figure 2 shows ML techniques for heart disease prediction.

3.2. Magnetocardiography-Based Ischemic Heart Disease Detection.
The framework of ischemic heart disease detection using ML classifiers is shown in Figure 3. The feature groups show that Pearson's coefficient was used to create a heatmap to investigate the association between characteristics and the target variable. Similar data points were grouped to evaluate clustering strength in order to
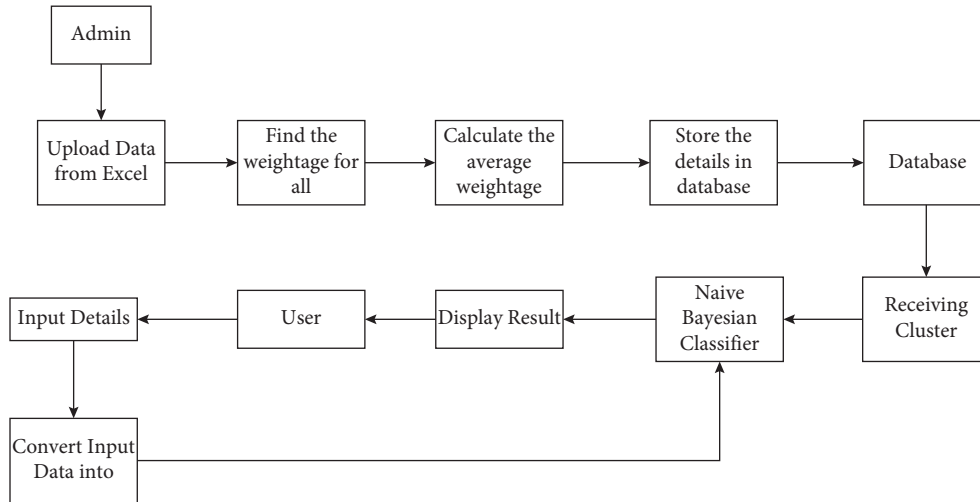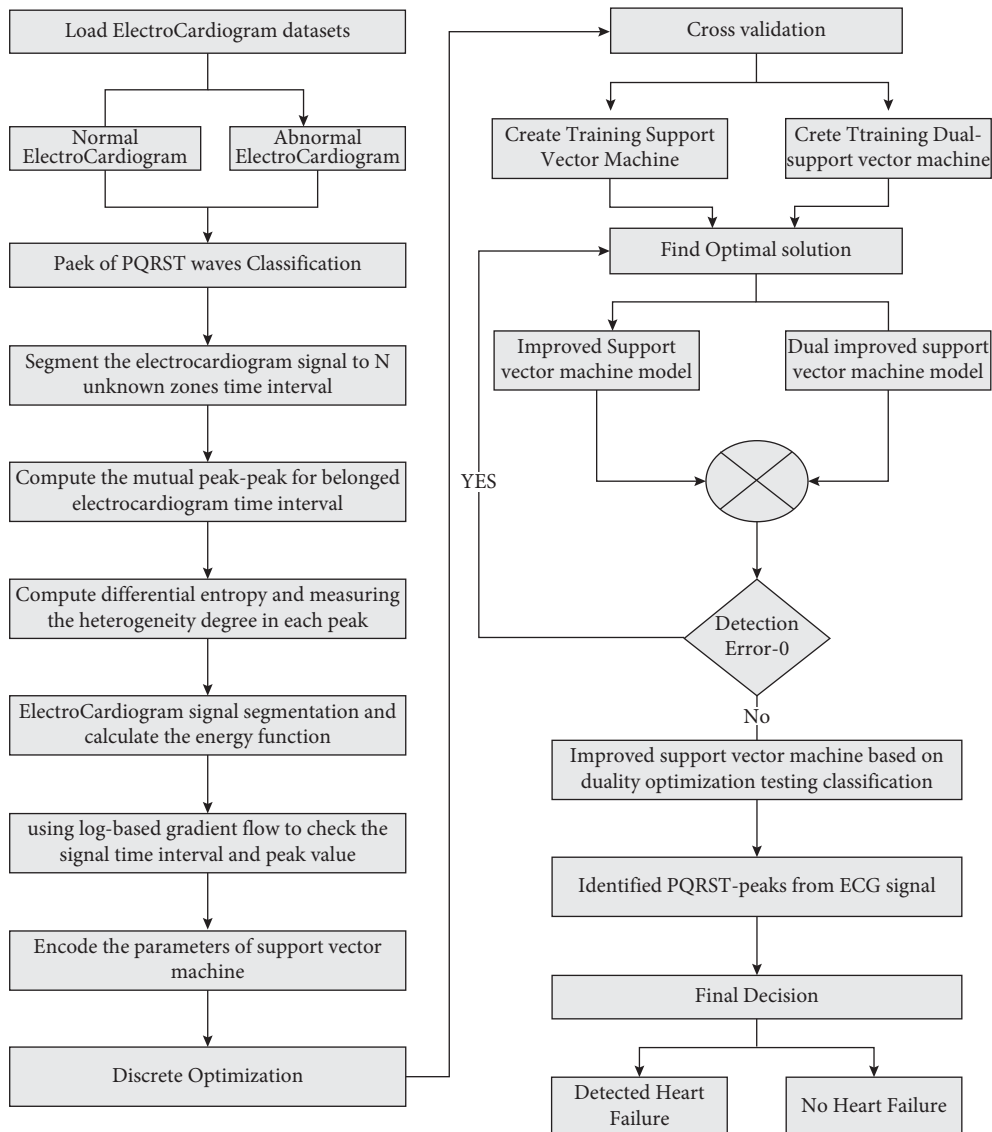
FIGURE 1: System architecture.

FIGURE 2: ML techniques for heart disease prediction.

Raw Data

Preprocessing with Feature extraction

Time   Frequency   Informations   Features

Classifier

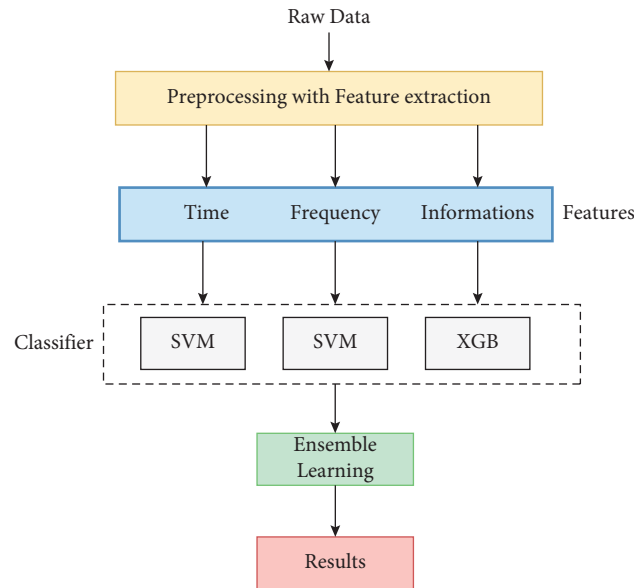SVM   SVM   XGB

Ensemble Learning

Results

FIGURE 3: Framework of ischemic heart disease detection.

investigate the links exhibited in the heatmap. With a mapping of the target variable, age and systolic blood pressure were grouped. This makes it easier to see the target variable's distribution. Preprocess the raw MCG signals for T-wave segmentation, filtering, and averaging [19, 20]. Then, 3 feature categories are extracted. Few of these features would be redundant, but still they are included in the classification procedure because past studies suggest that ML algorithms like SVM are not sensitive to redundant features' presence. On the other hand, omitting significant features is more hazardous than nonimportant features.

4. Feature Extraction

4.1. Time Domain. Time domain 18 features are summarized for describing the following:

(1) Maximum cardiac current characteristics

(2) Pattern of magnetic field map

(3) Distribution of negative/positive magnetic poles in the TT interval is independently distributed with a normal distribution that has a mean of 0 and constant variance

In addition, 6 related features of magnetic poles are presented to describe the following: change of negative/positive pole area in T wave and displacement of negative/positive in T wave. In the following study, it was determined that they were associated with the location of stenosis. Interesting characteristics of features from 2 aspects: dynamic changes and values at the peak of the T-wave during the interval TT. For achieving this, the interval TT is partitioned into 24 shorter subwindows, from which 18 time features are extracted.

4.2. Discrete Wavelet Transform Domain. A db-4 (Daubechies 4) DWT is applied with 4 levels on each of the 36

channels. The signal is decomposed into time-frequency components through DWT. The 4th level DWT low frequency part (A4 component, corresponding from 0 to 7.8 Hz approximately) is reconstructed utilizing IDWT (inverse DWT).

4.3. Information Theory. Based on the theory of electrophysiology, ischemic subjects' signal perturbation during the segment ST is higher than that of healthy persons. The 3 measures are selected from the following: (i) for each channel Shannon entropy, (ii) multidimensional Gini index, and (iii) entropy of SVD (singular value decomposition).

5. Heart Disease Prediction Using the XGBoost Algorithm

Develop HDPM for providing high prediction performance, absence/presence of heart disease, and to provide the present situation of a patient. The development process of HDPM is represented through flow chart which is shown in Figure 4. Firstly, collect the datasets of heart disease. Secondly, data preprocessing is carried out for the transformation of data. Third, apply the outlier detection based DBSCAN technique for determining the outlier data provided by optimal parameters [21]. Fourth, from the training dataset, remove the detected outlier data. Fifth, for balancing the training dataset, utilize SMOTE-ENN based data balancing technique. Sixth, to learn from the training dataset and for generating HDPM using MLA based on XGBoost. The main parameter with adjustment with $\max_{depth} = 5$, $\min_{childweight} = 1$, gamma $= 0$, subsample, $colsample_{bytree} = 0.8$, and $scale_{posweight} = 1$. Finally, for evaluating the presented model performance, performance metrics are presented. Within the CDSS, HDPM is generated [22, 23]. In this study, a 10-fold cross-validation technique is utilized for avoiding overfitting. The models are allowed by
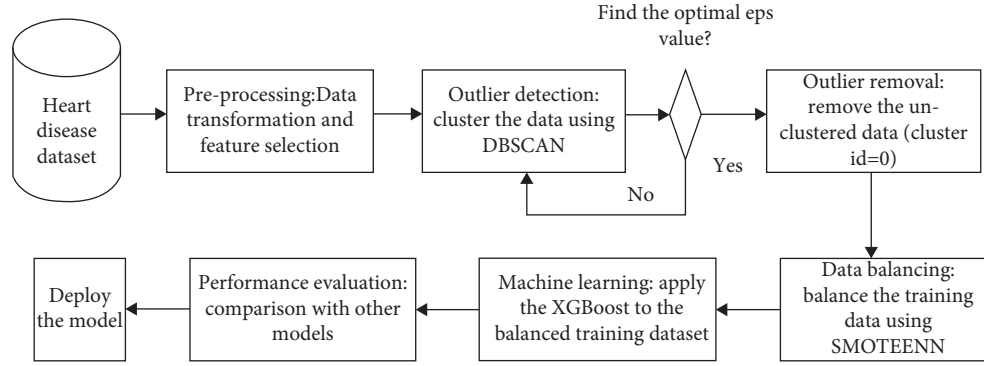
FIGURE 4: Proposed heart disease prediction model using XGBoost.

cross-validation to learn from various training datasets through repeated sampling. Therefore, data maximizing is utilized for validation, which helps for overfitting prevention. Past studies have demonstrated that the 10-cross field validation technique will be utilized for maintaining the bias variance trade-off that eventually provides a generalized model and protects against overfitting [24]. The proposed heart disease prediction model using XGBoost is shown in Figure 4.

6. Results

Different machine learning methods use different datasets with independent specification. Total 50 test cases are used in the prediction of heart diseases in the paper. Among these 50 test cases, 6 are false negatives, 1 is false positive, 18 are TPs, and 25 are TNs [25]. The collection of data is part of cardiovascular disease retrospective studies utilizing the recordings of multichannel MCG. There are 227 people with coronary stenosis and 347 people who are healthy in the database. There are 16 NSTEMI (non-ST-elevation myocardial infarction) instances in the sample. For the ischemic group, coronary angiography is performed. There were 227 IHD patients (left anterior descending) [26].

Several datasets are used to demonstrate the presented approach's capacity to detect HFD using ECG signals. The proposed method is tested on 38 real data recordings of ECG signals with HFD from the PhysioNet database. Consider the 5-section (P, Q, R, S, and T) segmentation issues of the ECG signal for each round [27]. It is also regarded as the three distinct zones (QRS, T, and P) that provide varying probabilities of peak-peak and all-waves time interval borderlines in the patient's ECG [21, 28, 29]. It has been discovered that three components of the processed ECG signal are segregated and distributed in the same way. For every patient with HFD, the ISVM-DO will identify all P, Q, R, S, and T wave peak values in the ECG waveform. As a result, the system extracts essential morphological aspects from processed ECG. Many random tests and cross-validation tests are used to train the proposed method predictor on 125 samples from the training dataset [30, 31].

For exploring how heart illness is recognized using an ML approach, two datasets of heart diseases (Cleveland and

Statlog; termed datasets I and II) are used. Although the original dataset comprises 79 raw attributes and 303 subjects, only 13 of them are utilized, and just one attribute is used as an output class. The remaining 297 subjects are used in the preprocessing step after 6 subjects were excluded from the dataset due to missing values. The XGBoost V0.81 Python library is used to implement XGBoost. Using the DBSCAN technique, remove outlier data from heart disease training datasets. The software XGBoost is used to generate HDPM [2, 32]. To evaluate the performance of ML techniques, 4 different parameters are utilized: recall, F1-measure, accuracy, and precision. For measuring the potential of these 4 parameters a confusion matrix is utilized from the model: F_n (false negative), T_n (true negative), F_p (false negative), and T_p (true positive). Number of subjects classified correctly as "positive" (heart disease presence) is known as T_p, number of subjects classified correctly as "negative" (absence/healthy heart disease) is represented as T_n. Similarly, number of subjects classified incorrectly as "negative" (when they have heart disease) is represented as F_n, and number of subjects classified incorrectly as "positive" when they not having heart disease is represented as F_p.

6.1. Accuracy. The ratio of accurately predicted predictions by the model to all types of completed predictions in the problem classification is known as accuracy.

$$\text{Accuracy} = \frac{\left(T_p + T_n\right)}{\left(T_p + T_n + F_p + F_n\right)}. \tag{1}$$

6.2. Precision. Precision or positive predictive is defined as the ratio of accurate positive scores (T_p) to the total number of positive scores $(T_p + F_p)$ predicted by the classification algorithm.

$$\text{Precision} = \frac{T_p}{\left(T_p + F_p\right)}. \tag{2}$$

6.3. Recall. The recall can be defined as the ratio of accurate T_p to the total T_p "+" F_n.

TABLE 1: Comparative analysis of different machine learning methods.

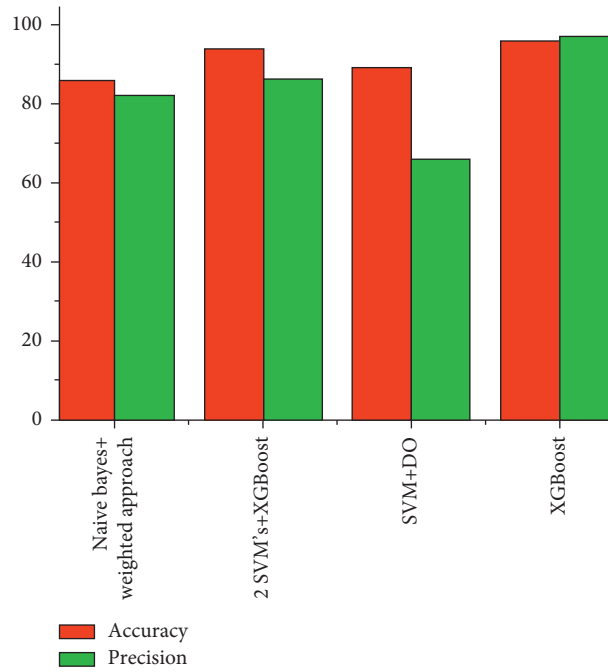| Methods	Accuracy	Precision	Recall	F1-measure
Naïve Bayes weighted approach	86.00	82.34	87.25	89.21
2 SVM's and XGBoost	94.03	86.56	94.78	92.79
SVM and DO	89.4	66.1	81.3	82.1
XGBoost	95.9	97.1	94.67	95.35

FIGURE 5: Different ML methods in terms of accuracy and precision.

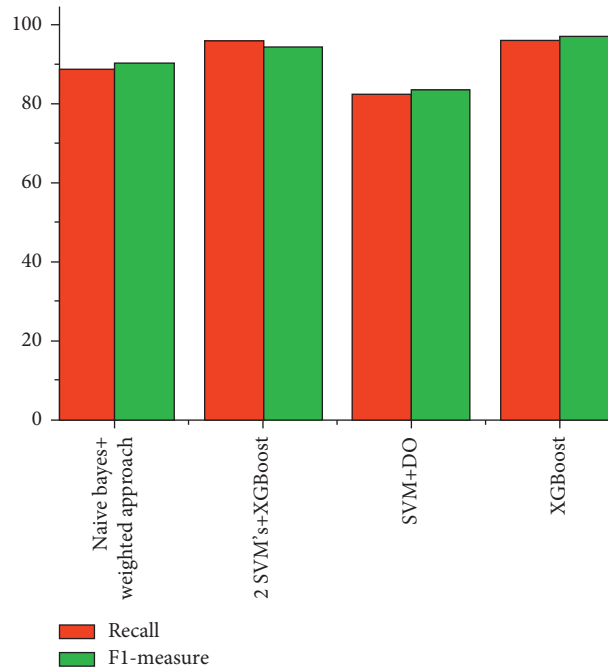

FIGURE 6: Different ML methods in terms of recall and F1-measure.

TABLE 2: Confusion metrics analysis for applying classifier.

Model name	Label	Predictive negative	Predictive positive
Naïve Bayes and weighted approach	Actual negative	6314	6214
	Actual positive	5076	5142
2 SVM's and XGBoost	Actual negative	6314	6014
	Actual positive	5410	5142
SVM and DO	Actual negative	5897	5001
	Actual positive	4517	4221
XGBoost	Actual negative	5794	5001
	Actual positive	4876	4221

$$\text{Recall} = \frac{T_p}{\left(T_p + F_n\right)}. \qquad (3)$$

6.4. F1-Measure. The F1-measure is a function of precision and recall. F1 must be 1 for the classification algorithm's good performance and 0 for bad performance.

$$\text{F1} - \text{measure} = 2 \times \frac{(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})}. \qquad (4)$$

Different ML techniques-based heart disease detection performance is evaluated with the different parameters, and these values are shown in Table 1.

The graphical representation of different ML techniques to predict heart disease in terms of accuracy and precision parameters is represented in Figure 5.

The graphical representation of the different ML techniques to predict heart disease in terms of recall and F1-measure parameters is represented in Figure 6.

From the abovementioned table and graphs, it is clear that the accuracy parameter is high in the XGBoost algorithm-based heart disease detection and low in the Naïve Bayes with weighted approach. Table 2 shows confusion metrics analysis for applying classifier.

7. Conclusion with Future Work

The survey on machine learning technology-based heart disease detection models is provided in this paper. Four approaches of ML models for heart disease detection are analyzed in this survey; these are the Naïve Bayes with weighted approach based prediction, 2 SVM's with XGBoost based prediction, an improved SVM (ISVM) based on duality optimization (DO) technique based prediction, and an XGBoost based prediction. According to the results analysis, the accuracy, precision, recall, and F1-measure parameters are high in the XGBoost algorithm-based heart disease detection and only accuracy is low for the Naïve Bayes with weighted approach than others, and the remaining precision, recall, and F1-measure values are low in SVM with duality optimization (DO) model. The present survey paper gives the best idea regarding different machine learning-based heart disease detection methods. This research can be updated in the future by adding more attributes to the heart disease dataset and making it more interactive for the users. It can also be carried out as a mobile application with reduced computing time and complexity. We will make changes to the system by linking it to the hospital's database.

Data Availability

The heart disease data that support the findings of this study are available on request from the corresponding author.

Conflicts of Interest

The authors of this manuscript declare that they have no conflicts of interest.

References

[1] R. Tao, S. Zhang, X. Huang et al., "Magnetocardiography-based ischemic heart disease detection and localization using machine learning methods," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 6, pp. 1658–1667, 2019.

[2] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, Article ID 81542, 2019.

[3] R. Spencer, F. Thabtah, N. Abdelhamid, and M. Thompson, "Exploring feature selection and classification methods for predicting heart disease," *Digital Health*, vol. 6, Article ID 2055207620914777, 2020.

[4] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDPM: an effective heart disease prediction model for a clinical decision support system," *IEEE Access*, vol. 8, Article ID 133034, 2020.

[5] I. D. Mienye and Y. Sun, "Improved heart disease prediction using particle swarm optimization based stacked sparse autoencoder," *Electronics*, vol. 10, no. 19, p. 2347, 2021.

[6] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection," *IEEE Access*, vol. 7, Article ID 180235, 2019.

[7] J. Wankhede, M. Kumar, and P. Sambandam, "Efficient heart disease prediction-based on optimal feature selection using DFCSS and classification by improved Elman-SFO," *IET Systems Biology*, vol. 14, no. 6, pp. 380–390, 2020.

[8] J. Wang, C. Liu, L. Li et al., "A stacking-based model for non-invasive detection of coronary heart disease," *IEEE Access*, vol. 8, Article ID 37124, 2020.

[9] Q. Zhenya and Z. Zhang, "A hybrid cost-sensitive ensemble for heart disease prediction," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 73, 2021.

[10] M. A. Khan and F. Algarni, "A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS," *IEEE Access*, vol. 8, Article ID 122259, 2020.

[11] F. I. Alarsan and M. Younes, "Analysis and classification of heart diseases using heartbeat features and machine learning algorithms," *Journal of Big Data*, vol. 6, no. 1, p. 81, 2019.

[12] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An automated diagnostic system for heart disease prediction based on ${\chi{2}}$ statistical model and optimally configured deep neural network," *IEEE Access*, vol. 7, Article ID 34938, 2019.

[13] X. Yuan, J. Chen, K. Zhang, Y. Wu, and T. Yang, "A stable ai-based binary and multiple class heart disease prediction model for IoMT," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 2032–2040, 2022.

[14] E. Sogancioglu, K. Murphy, E. Calli, E. T. Scholten, S. Schalekamp, and B. Van Ginneken, "Cardiomegaly detection on chest radiographs: segmentation versus classification," *IEEE Access*, vol. 8, Article ID 94631, 2020.

[15] F. Z. Abdeldjouad, M. Brahami, and N. Matta, "A hybrid approach for heart disease diagnosis and prediction using machine learning techniques," in *The Impact of Digital Technologies on Public Health in Developed and Developing Countries*, M. Jmaiel, M. Mokhtari, B. Abdulrazak, H. Aloulou, and S. Kallel, Eds., Springer International Publishing, Manhattan, NY, USA, 2020.

[16] A. Tyagi and R. Mehra, "Intellectual heartbeats classification model for diagnosis of heart disease from ECG signal using hybrid convolutional neural network with Goa," *SN Applied Sciences*, vol. 3, no. 2, p. 265, 2021.

[17] R. C. Ripan, I. H. Sarker, S. M. M. Hossain et al., "A data-driven heart disease prediction model through K-means clustering-based anomaly detection," *SN Computer Science*, vol. 2, no. 2, p. 112, 2021.

[18] S. U. Ghumbre and A. A. Ghatol, "Heart disease diagnosis using machine learning algorithm," in *Advances in Intelligent and Soft Computing*, S. C. Satapathy, P. S. Avadhani, and A. Abraham, Eds., Springer, Berlin, Heidelberg, 2012.

[19] L. C. Bergamasco, R. A. Oliveira, H. Wechsler, C. Dajuda, M. Delamaro, and F. L. Nunes, "Content-based image retrieval of 3D cardiac models to aid the diagnosis of congestive heart failure by using spectral clustering," in *Proceedings of the 2015 IEEE 28th International Symposium on Computer-Based Medical Systems*, pp. 183–186, Sao Carlos, Brazil, June 2015.

[20] A. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrès, "Classification models for heart disease prediction using feature selection and PCA," *Informatics in Medicine Unlocked*, vol. 19, Article ID 100330, 2020.

[21] J. Nourmohammadi-Khiarak, M.-R. Feizi-Derakhshi, K. Behrouzi, S. Mazaheri, Y. Zamani-Harghalani, and R. M. Tayebi, "New hybrid method for heart disease diagnosis utilizing optimization algorithm in feature selection," *Health Technology*, vol. 10, no. 3, pp. 667–678, 2020.

[22] L. Verma, S. Srivastava, and P. C. Negi, "An intelligent noninvasive model for coronary artery disease detection," *Complex & Intelligent Systems*, vol. 4, no. 1, pp. 11–18, 2018.

[23] A. K. Pathak and J. Arul Valan, "A predictive model for heart disease diagnosis using fuzzy logic and decision tree," in *Smart Computing Paradigms: New Progresses and Challenges*, A. Elçi, P. K. Sa, C. N. Modi, G. Olague, M. N. Sahoo, and S. Bakshi, Eds., Springer, Singapore, 2020.

[24] J. Kaur and B. S. Khehra, "Fuzzy logic and hybrid based approaches for the risk of heart disease detection: state-of-the-art review," *Journal of the Institution of Engineers: Series B*, 2021.

[25] V. Gomathy, N. Padhy, D. Samanta, M. Sivaram, V. Jain, and I. S. Amiri, "Malicious node detection using heterogeneous cluster based secure routing protocol (HCBS) in wireless adhoc sensor networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 11, pp. 4995–5001, 2020.

[26] R. Thanga Selvi and I. Muthulakshmi, "An optimal artificial neural network based big data application for heart disease diagnosis and classification model," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 6, pp. 6129–6139, 2021.

[27] A. K. Biswal, D. Singh, B. K. Pattanayak, D. Samanta, S. A. Chaudhry, and A. Irshad, "Adaptive fault-tolerant system and optimal power allocation for smart vehicles in smart cities using controller area network," *Security and Communication Networks*, vol. 2021, Article ID e2147958, 2021.

[28] A.-Z. Sultan Bin Habib, T. Tasnim, and M. M. Billah, "A study on coronary disease prediction using boosting-based ensemble machine learning approaches," in *Proceedings of the 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET)*, pp. 1–6, Dhaka, Bangladesh, December 2019.

[29] V. Shorewala, "Early detection of coronary heart disease using ensemble techniques," *Informatics in Medicine Unlocked*, vol. 26, Article ID 100655, 2021.

[30] A. Guo, M. Pasque, F. Loh, D. L. Mann, and P. R. O. Payne, "Heart failure diagnosis, readmission, and mortality prediction using machine learning and artificial intelligence models," *Current Epidemiology Reports*, vol. 7, no. 4, pp. 212–219, 2020.

[31] P. Sivakumar, R. Nagaraju, D. Samanta, M. Sivaram, M. N. Hindia, and I. S. Amiri, "A novel free space communication system using nonlinear InGaAsP microsystem resonators for enabling power-control toward smart cities," *Wireless Networks*, vol. 26, no. 4, pp. 2317–2328, 2020.

[32] B. A. Tama, S. Im, and S. Lee, "Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble," *BioMed Research International*, vol. 2020, Article ID 9816142, 2020.

Hindawi

Research Article

Identification and Prediction of Chronic Diseases Using Machine Learning Approach

Rayan Alanazi ⬤

Department of Computer Science, College of Science and Arts in Qurayyat, Jouf University, Sakakah, Saudi Arabia

Correspondence should be addressed to Rayan Alanazi; rmalanazi@ju.edu.sa

Nowadays, humans face various diseases due to the current environmental condition and their living habits. The identification and prediction of such diseases at their earlier stages are much important, so as to prevent the extremity of it. It is difficult for doctors to manually identify the diseases accurately most of the time. The goal of this paper is to identify and predict the patients with more common chronic illnesses. This could be achieved by using a cutting-edge machine learning technique to ensure that this categorization reliably identifies persons with chronic diseases. The prediction of diseases is also a challenging task. Hence, data mining plays a critical role in disease prediction. The proposed system offers a broad disease prognosis based on patient's symptoms by using the machine learning algorithms such as convolutional neural network (CNN) for automatic feature extraction and disease prediction and K-nearest neighbor (KNN) for distance calculation to find the exact match in the data set and the final disease prediction outcome. A collection of disease symptoms has been performed for the preparation of the data set along with the person's living habits, and details related to doctor consultations are taken into account in this general disease prediction. Finally, a comparative study of the proposed system with various algorithms such as Naïve Bayes, decision tree, and logistic regression has been demonstrated in this paper.

1. Introduction

All over the world, chronic diseases are a critical issue in the healthcare domain. According to the medical statement, due to chronic diseases, the death rate of humans increases. The treatments given for this disease consume over 70% of the patient's income. Hence, it is highly essential to minimize the patient's risk factor that leads to death. The advancement in medical research makes health-related data collection easier [1, 2]. The healthcare data includes the demographics, medical analysis reports, and the history of disease of the patient. The diseases caused could be varied based on the regions and the living habitats in that region. Hence, along with the disease data, the environmental condition and the living habitat of the patient should also be recorded in the data set.

In recent years, the healthcare domain is evolving more due to the integration of information technology (IT) in it. The intention to integrate IT in healthcare is to make the life

of an individual more affordable with comfort as smartphones made one's life easier [3]. This could be possible by making healthcare to be intelligent, for instance, the invention of the smart ambulance, smart hospital facilities, and so on, which helps the patients and doctors in several ways [4]. The research on a specified region for patients affected by chronic diseases every year had been held and found that the difference between the patients in genderwise is very small, and it is found that the large number of patients were admitted in the year 2014 for treating chronic diseases. The use of structured and unstructured data provides highly accurate results instead of using only structured data. Since the unstructured data includes the doctor's records on the patients related to diseases and the patient's symptoms and grievances faced by them, explained by themselves, which is an added advantage when used along with the structured data that consists of the patient demographics, disease details, living habitats, and laboratory test results [5, 6]. It is difficult to diagnose rare diseases. Hence, the use of self-

reported behavioral data helps differentiate the individuals with rare diseases from the ones with common chronic diseases. By using machine learning approaches along with questionnaires, it is believed that the identification of rare diseases is highly possible [7].

In the last decade, some innovative technologies had been introduced to rapidly collect the data such as MRI (magnetic resonant imaging) readouts, ultrasonography, social media gained data, and electronically gained activity, behavioral, and clinical data. These big data sets of healthcare are high-dimensional, which means the number of features recorder per observation might be greater than the total observations. They are noisy, sparse, cross-sectional, and lacks statistical power. By applying machine learning techniques, the issues in the high-dimensional data sets can be overcome [8]. Machine learning contributes more in several domains. Many of the complex models make use of exiting larger training data, simultaneously at the edge of a major shift in healthcare epidemiology [9]. These data can enhance the knowledge gain in the risk factors of diseases to reduce healthcare-associated infections, improve patient risk stratification, and find the way of transmitting the infectious diseases [10]. Machine learning can facilitate the analysis of laboratory results and other details of patients for the early detection of diseases. The low-level data could be converted to high-level knowledge via knowledge discovery in the database so as to gain knowledge about the disease patterns to support early detection [11]. The data collected for creating a data set should be preprocessed for its missing values, and then only the important features needed for accurate disease prediction are selected so as to enhance the prediction accuracy and minimizing the time taken for model training [12].

In the era of the Internet and technologies, people are not concerned about their health and lives. As everyone is interested in surfing and social media activities, they ignore visiting hospitals for their health checkup. By taking this activity as an advantage, a machine learning model that takes the symptoms given as input and predicts the possibility and risk of the disease affected or the development of such diseases in an individual should be developed [13, 14]. The more common chronic diseases are diabetes, cardiovascular diseases, cancer, strokes, hepatitis C, and arthritis. As these diseases persist for a long time and have a high mortality rate, the diagnosis of such diseases is highly important in the healthcare domain. Foreseeing the disease can help take preventive actions and avoid getting affected by it, and early detection of it can help provide better treatment [15]. There are various techniques in machine learning such as supervised, semisupervised, unsupervised, reinforcement, evolutionary, and deep learning. The problem is associated with the processing of extracted features from real data and structured as vectors [16]. The processing quality is based on the proper combination of those vectors. But, most of the times, the high dimensionality of the vectors or the discrepancies in the data makes a big issue. Hence, it is important to reduce the dimensionality of the data set even if it leads to a little loss of details to make the data set a highly compatible dimension. This reduction in the dimensionality of the data set improves the model performance [17].

The system of chronic diseases management is essential for those affected by such diseases and in need of proper medical assessment and treatment information [18]. Also, this system can be useful for individuals who are in need of self-care to improve their health condition, since it is proved that self-management is the primary care of those with chronic diseases, and it is considered as the unavoidable part of treatment. With the use of mobile applications, the health information of patients can be recorded, and they serve as a better tool to enable self-management [19]. To effectively predict a disease, information such as narration about the symptoms felt by the patients, details of consultation with medical practitioners, lab examination results, and computed tomography and X-ray images [20]. Little research is performed in identifying the accuracy and predictive power for developing a machine learning model with only information from lab examination results for the diagnosis of diseases. And, for performance enhancement, ensemble machine learning and deep learning model can be used [21, 22]. In the healthcare domain, artificial intelligence (AI) plays a major role in automating the roles involved in disease diagnosis and treatment suggestions and also schedules perfect timing by the medical practitioners to perform various obligations that cannot be automated [23].

The major objective of the proposed system is to identify and predict chronic disease in an individual using a machine learning approach [24, 25]. The data set comprises both the structured data that includes the patient's age, gender, height, weight, and so on, excluding the patient's personal information such as name and ID, and the unstructured data that includes the patient's symptoms, information related to consultation about the disease with the doctors, and the living habits of that individual [26]. These data are preprocessed for finding the missing values. They are then reconstructed to increase the quality of the model, thereby increasing the prediction accuracy. For prediction, the machine learning algorithms such as CNN and KNN are used [27, 28]. This paper is organized as the details of the related works carried out while doing the research are given in Section 2, the preliminaries of the algorithms used in given in Section 3, the description of the proposed methodology in Section 4, the result and discussion part are given in Section 5 followed by the conclusion in Section 6, and finally, a list of references used in this study has been given.

2. Related Work

This section describes the related works that are performed in developing the proposed model for predicting chronic diseases. The following are the discussions made by reviewing the existing literature that helps develop the proposed system efficiently and effectively.

The objective variable of the study in [29] is the resource consumption such as medical and long-term care expenses and a predictive model for medical care using a random forest machine learning algorithm [30]. This method uses data of more than 100 pieces that includes preventive activities, clinical tests, and medical practices. This model uses mean decrease Gini for classification and for regression

mean square error (MSE) is used [31, 32]. The training model uses a grid search for hyperparameter tuning and is validated using K-fold cross-validation. Along with the objective variable, exploratory variables such as age, gender, and analysis period are also included, since the aim of this paper is proper management of the budget for medical care [33]. A review that highlights the applications of machine learning techniques in various medical practices such as predicting, diagnosing, and prognosis of diseases such as multiple sclerosis, autoimmune chronic kidney disease, autoimmune rheumatic disease, and inflammatory bowel disease and for the selection of treatments and stratification of patients; drug development; drug repurposing; target interpretation; and validation has been given in [34, 35]. This paper also provides a detailed description of the challenges faced by the machine learning approaches such as the need for quality data in preparation of robust models, external model validation using the independent data set, difficulties faced during implementation of a model, and ethical concerns. A predictive model for chronic kidney disease is explained in [36, 37]. This model is developed using four machine learning approaches such as support vector machine (SVM), logistic regression (LR), decision tree (DT), and KNN for classification purposes. The data set used in this paper is the Indian chronic kidney disease (CKD) that consists of 400 occurrences, 24 features, and 2 classes obtained from the UCI machine learning repository. The developed model is evaluated using a 5-fold cross-validation process, and the experiment is conducted on the Weka data mining tool and MATLAB and finally concluded that the SVM classifier attains higher accuracy when compared to the others.

A system that can predict multiple diseases with the help of various machine learning algorithms such as Naïve Bayes, KNN, DT, random forest, and SVM algorithms has been described in [38] to bridge the gap among the patients and the doctors to achieve their own goals. The existing approaches in the field of automatic disease prediction lack the patient's trust in the model's prediction and also reduce the need for doctors, which makes the doctors get panic about their livelihood. But this method integrates a module for doctor recommendation that solves both the issues by making sure the patient to trust due to the intervention of doctors and also improves the business of doctors. A model called PARAMO, which is a platform of a parallel predictive model that uses electronic health records (EHR) for healthcare analysis, has been implemented in [39]. This method comprises three phases, namely, the generation of the dependency graph, which removes redundancy and identifies dependency; then execution engine for dependency graph, which includes prioritizing, scheduling, and parallel execution; and finally the parallelization infrastructure. The PARAMO model is tested with three sets of real data, that is, small, medium, and large data sets that includes the medications, diagnosis data, and lab records, obtained from EHR that ranges from 5,000 to around 300,000 patients. In addition to this, the small and large sets include the procedure data, and the medium set includes the symptoms of heart failure that are taken from medical records [40]. An efficient recommendation system for chronic disease diagnosis has been demonstrated in [41]. This method uses a data mining approach. The data set used in this system includes medical data and two-dimensional data. The medical data include the data obtained from sensors or medical data entries, and the two-dimensional data include the external user and the item features. For enhancing the accuracy of prediction, the decision tree approach, which is a highly prevalent data mining approach, is used for classification. Various decision tree classifiers such as random forest, REP tree, decision stump, and J48 are involved in the creation of this predictive model. This model is tested with randomly selected 20 samples and found that the RF outperforms the other three algorithms.

Prediction of 3 types of immune diseases such as allergy, infectious, and autoimmune diseases using decision tree, maximum margin learning, and instance-based learning, respectively, has been given in [42]. The correlation between the classification of immunogens and its physicochemical properties is one of the purposes of this study. The immunogen data such as the stats of diseases, responses from B-cell, discontinuous epitope location, host, source organisms, and so on are collected from Immune Epitope Database (IEDB) and analyzed its 6 physicochemical properties such as PSSM (position-specific scoring matrix) information per position, hydrophilic scale, flexibility, antigenic propensity, hydropathy index, and side chain polarity. This system is tested using a method called leave-one-out cross-validation for the performance of prediction outcomes with parameters such as accuracy and F-score. A risk prediction model for predicting disease risks using a random forest machine learning approach from highly imbalance data has been described in [43]. The data set used in this approach is the Nationwide Inpatient Sample (NIS), which includes 8 million records of hospital stays with 126 clinical as well as nonclinical data. The nonclinical data comprises patient's demographics, hospital location, date and year of admission, pin code, treatment/diagnosis cost, and duration of stay in a hospital ward. The clinical data comprises the treatment procedures, its categories, diagnosis categories, and its codes. Each record has a vector containing 15 diagnosis codes characterized by International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM). As the unbalance data produces undesirable results, a repeated random sampling method is employed to solve this issue. The developed model is evaluated using SVM, RF ensemble learning, bagging, and boosting algorithms. The study [44] demonstrates a novel adaptive probabilistic divergence-based feature selection algorithm to predict chronic kidney disease in its earlier stage. This algorithm is based on statistical and divergence information theory. For classification, the hyperparameterized logistic regression model is used in this study. The data set used in this approach is obtained from various hospitals and laboratories with information of 630 patients with 52 attributes, and this data set is given to the physician for verification of its correctness. The model developed is evaluated using the data sets of diabetes, heart, and kidney diseases, and the performance evaluation metrics

followed in this study is the precision, recall, F1-score, and ROC (receiver operating characteristics) curve.

A system that enhances the risk prediction of a patient's health condition using a deep learning approach on big data and a revised fusion node model has been demonstrated in [45]. This deep learning model for extracting the data and logical inference is made of the combination of complex machine learning algorithm such as Bayesian fusion and neural networks. The architecture of this system consists of five layers, namely, the data layer that is responsible for data collection, data aggregation layer for data acquisition from several data sources and desired format changing, analytics layer to do proper analytics on the data aggregated, information exploration layer to create the output that makes the results of analytics understandable for users, and big data governance layer that is responsible for managing the above layers. Also, in this paper, the application of MapReduce is discussed for optimizing the analytics efficiency and also inspires the design of SOA (service-oriented architecture) for making the external systems easily access the results from analytics. A machine learning model of disease prediction cost has been implemented in [46] that uses big data, which includes structured and unstructured data for preparing the data set and the developed model is made available at affordable. The prediction algorithm used in this method is the decision tree algorithm and the MapReduce algorithm is applied for enhancing the efficiency of the operation. The advantages of this model are reduction in retrieval time of queries, improved accuracy. A method of predicting the risk of chronic kidney disease using zub machine learning approaches has been described in [47, 48]. Two types of data sets are used in this method. One is from UCI with 400 instances and 35 features, and the other is a real-time data set obtained from Khulna City Medical College with 55 instances and 25 features. Data processing is done using Pandas and Numpy libraries, and the missing data are handled using median filtering. Feature extraction is performed using the Chi-square test. Model evaluation is performed using 10-fold cross-validation. Artificial neural network (ANN) and random forest algorithms are used for disease classification. This method is believed that it can predict the risk of chronic kidney disease in its earlier stage [49–52].

3. Preliminaries

3.1. Chronic Disease.
According to US National Center for Health Statistics, chronic diseases are diseases that last for a long period of time, that is, more than three months. These diseases are neither treated by medicines nor prevented by vaccines. The major cause of chronic diseases is the use of tobacco, unhealthy food habits, and lack of physical activity. Also, this disease can commonly be caused due to ageing. Chronic diseases include cardiovascular disease, cancer, arthritis, diabetes, obesity, epilepsy and seizures, and problems in oral health [35].

Cardiovascular disease includes heart disease and stroke, which highly lead to death. This disease is caused due to the use of tobacco, intake of nutritionless food, and lack of physical activity. When these activities are changed by the patient, they might have the chance to reduce the impact on controlling and preventing cardiovascular disease.

Next to cardiovascular disease, cancer such as colon cancer and breast cancer is considered the deadliest disease. It can be controlled only by prevention, early detection, and proper medical support. Minimizing the prevalence of environmental and behavioral factors that causes cancer reduces the chance risk of causing it.

The chronic disease such as arthritis causes inflammation in the joints, causes pain, and stiffness that increases due to ageing. There is an availability of cost-effective methods for reducing the effects caused by arthritis but are not used much. The effects of arthritis can be reduced by following moderate exercises regularly.

Diabetes is a serious and high-money-consuming disease. The impact of diabetes can be reduced by self-care and early detection of the disease [53]. Around 7 million people over the age of 65 or above are affected by this disease particularly type 2 diabetes.

Since 1980, obesity is more common in adults for all age groups. The one who is overweight or obese can develop the risk of getting high blood pressure (BP), heart diseases, diabetes, and arthritis. Obesity can also cause some types of cancers.

Epilepsy and seizures are highly costly in treatment [54]. This disease is common among all age groups, especially in young and elders.

Oral health problems are a crucial issue that attains special attention in the health of older people. This is a serious issue, since it affects the normal day-to-day actions of a person such as speak, chew, swallow, and maintain a nutritional food plan.

3.2. Convolutional Neural Network (CNN).
The ConvNet or CNN is an algorithm of deep learning that fetches the input and assigns the bias and weights to its several aspects and then distinguishes one from the other [55] as shown in Algorithm 1. The major reason for using CNN is that it requires only few efforts in preprocessing the data when compared with other algorithms, since the CNN can learn to optimize the filters through automate learning [56]. The output layer of CNN can be calculated using the following expression:

$$\text{size of output layer} = \text{input size} - (\text{filter size} - 1). \quad (1)$$

3.3. K-Nearest Neighbor (KNN).
KNN is a supervised machine learning algorithm, which analyzes the similarities between the new data and the existing data and adds the new data into the category that is highly similar to the available categories [57] as shown in Algorithm 2. The KNN can be used in classification as well as regression tasks, but it is most commonly used in classification. This algorithm is also called the lazy learner algorithm; since it will not learn instantly from the training data, it stores the data set and does its action during the classification process. The calculation of Euclidean distance is expressed mathematically as follows:

$$x^2 = c - a^2 + d - b^2. \quad (2)$$

Consider the activation function as,
lambda x: $1.0/(1.0 + np.\exp(-x))$
#Sigmoid function
Input: np.random.randn [3 1]
hiddenLayer-1 = activation (np.dot (W_1, input) + b_1)
hiddenLayer-2 = activation (np.dot (W_2,
    hiddenLayer-1) + b_2)
Output: np.dot (W_3, hiddenLayer-2) + b_3
were, $W_1, W_2, W_3, b_1, b_2, b_3$ are learnable parameters

ALGORITHM 1: Convolutional neural network algorithm.

Input: x, S, d
Output: Class of X
for $(X', l') \in S$ **do**
Compute the distance $d(X', X)$
end for
Sort the distance $|S|$ in ascending order
Count the number of occurrences of each class l_j
    among the k-nearest neighbors
Assign the most frequent class to X

ALGORITHM 2: K-nearest neighbor algorithm.

4. Proposed Methodology

In this section, a detailed description of the data set creation, model preparation, and disease prediction has been given. The first action is data collection. Our proposed system collects structured and unstructured data obtained from various sources. After data collection, they are subjected to preprocessing and are split into cleaning and test data sets. Then the training data set is trained with the machine learning algorithms such as CNN and KNN to a number of epochs for improving the accuracy of the prediction results. After multiple epochs, once the desired target is achieved, the developed model is ready for testing.

At this step, the model is tested with the test data set to verify the model performance with brand-new data that were not used for training. If the model attains the desired accuracy in test data, then the proposed model is ready for deployment as shown in Figure 1.

4.1. Data Collection.
The real-life data that includes structured data such as patient basic information including demographics, living habitat, and lab test results and the unstructured data such as the symptoms of the disease faced by the patient and their consultation with the doctor. The data set excludes the patient's personal details such as name, ID, and location so as to preserve their privacy.

4.2. Preprocecssing.
The collected data are preprocessed for the availability of missing values in most of the structured data. Hence, it is essential to fill out the missed data or remove or modify them to enhance the quality of the data
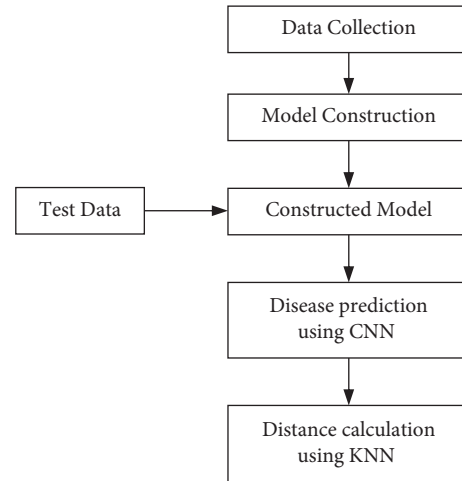

FIGURE 1: Architecture of proposed disease and risk prediction system.

set. The preprocessing step also eliminates the commas, punctuations, and white spaces. Once the preprocessing of data has been completed, it is then subjected to feature extraction followed by disease prediction.

4.3. Model Description.
As discussed above, the data set consists of both structured and unstructured data. The structured data comprises patient demographics and the data related to the cause for the disease such as age, gender height, weight, and so on, patient's living habitat, laboratory test results, and the disease that they are affected in tabular format. The unstructured data comprises patient's disease symptoms and the information about the interrogation with doctors in text format. The unstructured data is an added advantage of the prediction task to get a more accurate results. The data set is split into 80% for training and 20% for testing.

4.4. Disease Prediction Using CNN.
The proposed system uses the CNN algorithm in the prediction of chronic disease. At first, the data set is converted into vector form, followed by word embedding to adopt zero values for filling the data. It is then given to the convolution layer.

The pooling layer takes the input from the convolution layer and follows the max pooling operation. The output of max pooling is given to the fully connected layer, and then finally, the output layer provides the classification results. Figure 2 shows the block diagram of the convolutional neural network.

4.5. Distance Calculation Using KNN.
In K-Nearest Neighbor (KNN), the value of K is known, and the features that are similar to the K value are called the nearest neighbor. The nearest neighbor to the known K value is chosen, and the nearest distance between them is calculated. The feature with less distance value is considered to be the exact match, which is the final disease prediction output. In the proposed
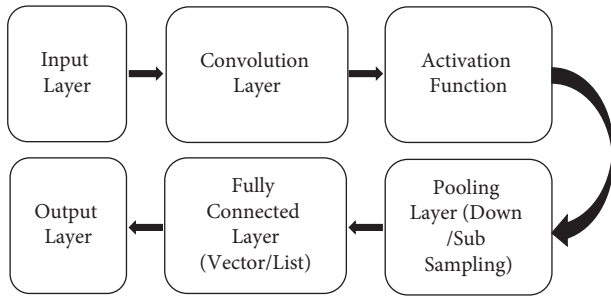
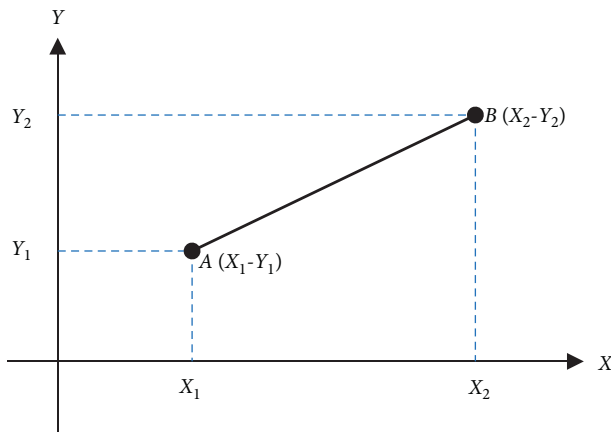FIGURE 2: Block diagram of convolutional neural network.

FIGURE 3: Calculation of Euclidean distance.

system, Euclidean distance is used, since the result obtained by it is better when compared to other distance calculation methods. It is a nonparametric algorithm since it will not take decisions on original data. In KNN, the training input data are located in X and Y axes, and the test data are located in the plots of X and Y axes. Then, the plots of test data with less distance are chosen and are considered as the desired target. It is important to choose the value of the nearest K point should be always odd.

The calculation of Euclidean distance can be performed by using the following formula and is represented in Figure 3:

$$D = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \cdots + (X_n - Y_n)^2}. \qquad (3)$$

5. Performance Evaluation

For evaluating the proposed disease prediction model, four performance evaluation metrics are used. The confusion matrix consists of the true positives (TP), which is the correct prediction of the target as a patient with chronic disease; the true negatives (TN), which is the correct prediction of the persons without diseases; false positives (FP), which is the incorrect prediction of the healthy person as a diseased person, and false negatives (FN), which is the incorrect prediction of the target as healthy persons. The following is the description of the four performance evaluation parameters.

5.1. Accuracy.
The classification accuracy is described as the ratio of correct predicted values to the total predicted values and is depicted mathematically as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} * 100. \qquad (4)$$

5.2. Precision.
The precision or positive predictive value (PPV) is described as the ratio of correct prediction to the total correct values including the true and false predictions and is depicted mathematically as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \qquad (5)$$

5.3. Recall.
The recall or sensitivity or true positive rate (TPR) is described as the ratio of correct predicted values to the sum of correct positive predictions and the incorrect negative predicted values and is depicted mathematically as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \qquad (6)$$

5.4. F1-Score.
The F-measure (F_β) is described as the weighted average of the values obtained from the calculation of precision and recall parameters. Whenever the distribution of class is not even, then the value of $F_1 - \text{Score}$ is highly important than the accuracy value. And whenever the values of false positives and negatives are dissimilar, the value of $F_1 - \text{Score}$ is highly suitable. The $F_1 - \text{Score}$ is depicted mathematically as follows:

$$F_\beta = \frac{\left(1 + \beta^2\right)(\text{Precision} * \text{Recall})}{\left(\beta^2 * (\text{Precision} + \text{Recall})\right)}. \qquad (7)$$

By simplifying using $\beta = 1$,

$$F_1 - \text{Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \qquad (8)$$

The obtained values of precision, recall, and F1-score of the proposed CNN and KNN model is compared with the values of the performance metrics of Naïve Bayes, decision tree, and logistic regression algorithms, and the results are tabulated in Table 1.

The accuracy is the important parameter since the prediction result is the important factor for the patient, and if it is wrong, then it will be a detriment to them. The other parameters such as precision, recall, and F1-score are for the evaluation of the model performance as shown in Table 1.

Figure 4 shows the graphical representation of the comparison results of accuracies of the proposed and other algorithms. This graph illustrates the variations in the prediction accuracies of the four algorithms such as the Naïve Bayes, decision tree, logistic regression, and the proposed CNN and KNN algorithms as 52%, 62%, 86%, and 96%, respectively. This shows that the proposed system achieves the highest accuracy of 96% when compared to the other machine learning algorithms.

TABLE 1: Performance evaluation comparison.

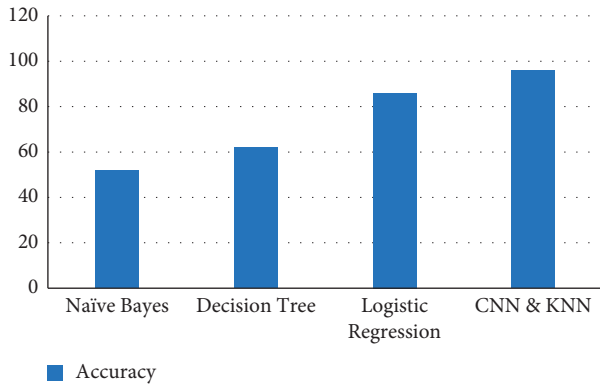|	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Naïve Bayes	52	52	80	65
Decision tree	62	64	60	62
Logistic regression	86	84	88	82
CNN and KNN	96	93	99	97

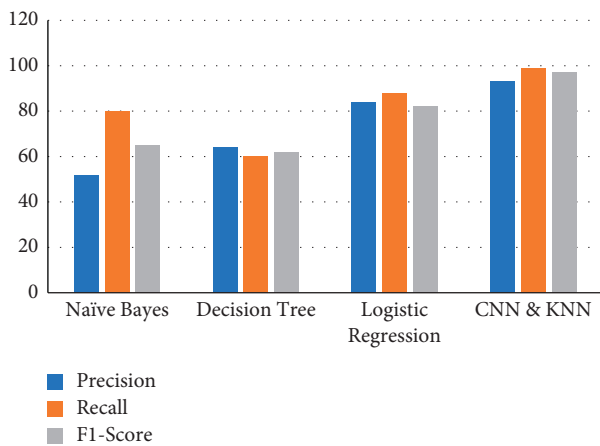FIGURE 4: Comparison of accuracies of proposed and other algorithms.

FIGURE 5: Comparison of other performance evaluation metrics of proposed and other algorithms.

Figure 5 shows the graphical representation of the comparison precision, recall, and F1-score values of the proposed and other algorithms. This graph illustrates the variations in the three performance evaluation parameters of the four algorithms such as the Naïve Bayes, decision tree, logistic regression, and the proposed CNN and KNN algorithms as 52%, 64%, 84%, and 93%, respectively, for precision; 80%, 605, 88%, and 99%, respectively, for recall; and 65%, 62%, 82%, and 97%, respectively, for F1-score. These results shows that the prosed model developed using CNN and KNN algorithm is considered to be the best of the remaining three algorithms with 93%, 99%, and 97% for precision, recall, and F1-score, respectively, which is higher when compared to the others.

6. Conclusion

This paper proposed a method of identification and prediction of the presence of chronic disease in an individual using the machine learning algorithms such as CNN and KNN. The advantage of the proposed system is the use of both structured and unstructured data from real life for data set preparation, which lacks in many of the existing approaches. In this paper, the performance of the proposed model is compared with other algorithms such as Naïve Bayes, decision tree, and logistic regression algorithms. The results show that the proposed system provides an accuracy of 95% that is higher than that of the other two algorithms. It is highly believed that the proposed system can reduce the risk of chronic diseases by diagnosing them earlier and also reduces the cost for diagnosis, treatment, and doctor consultation.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

References

[1] G. Battineni, G. G. Sagaro, N. Chinatalapudi, and F Amenta, "Applications of machine learning predictive models in the chronic disease diagnosis," *Journal of Personalized Medicine*, vol. 10, no. 2, p. 21, 2020.

[2] B. Manjulatha and P. Suresh, "An ensemble model for predicting chronic diseases using machine learning algorithms," in *Smart Computing Techniques and Applications*, pp. 337–345, Springer, New York, NY, USA, 2021.

[3] C.-H. Jen, C.-C. Wang, B. C. Jiang, Y.-H. Chu, and M.-S. Chen, "Application of classification techniques on development an early-warning system for chronic illnesses," *Expert Systems with Applications*, vol. 39, no. 10, pp. 8852–8858, 2012.

[4] D. Gupta, S. Khare, and A. Aggarwal, "A method to predict diagnostic codes for chronic diseases using machine learning techniques," in *Proceedings of the 2016 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 281–287, IEEE, Greater Noida, India, April 2016.

[5] M. Chen, H. Yixue, K. Hwang, L Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *Ieee Access*, vol. 5, pp. 8869–8879, 2017.

[6] R. Ge, R. Zhang, and P Wang, "Prediction of chronic diseases with multi-label neural network," *IEEE Access*, vol. 8, pp. 138210–138216, 2020.

[7] H. MacLeod, S. Yang, O. Kim, C. Kay, and S. Natarajan, "Identifying rare diseases from behavioural data: a machine learning approach," in *Proceedings of the 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 130–139, IEEE, Washington, DC, USA, June 2016.

[8] M. A. Myszczynska, P. N. Ojamies, A. M. B. Lacoste et al., "Applications of machine learning to diagnosis and treatment of neurodegenerative diseases," *Nature Reviews Neurology*, vol. 16, no. 8, pp. 440–456, 2020.

[9] I. Preethi and K. Dharmarajan, "Diagnosis of chronic disease in a predictive model using machine learning algorithm," in *Proceedings of the 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICST-CEE)*, pp. 191–96, IEEE, Bengaluru, India, October 2020.

[10] J. Wiens and E. S Shenoy, "Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology," *Clinical Infectious Diseases*, vol. 66, no. 1, pp. 149–153, 2018.

[11] S. Swaminathan, K. Qirko, T. Smith et al., "A machine learning approach to triaging patients with chronic obstructive pulmonary disease," *PLoS One*, vol. 12, no. 11, Article ID e0188532, 2017.

[12] Z. Wang, J. W. Chung, X. Jiang, Y. Cui, M. Wang, and A. Zheng, "Machine learning-based prediction system for chronic kidney disease using associative classification technique," *International Journal of Engineering & Technology*, vol. 7, pp. 1161–1167, 2018.

[13] A. Kumar and A. Pathak, "A machine learning model for early prediction of multiple diseases to cure lives," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 6, pp. 4013–4023, 2021.

[14] C. Kalaiselvi, "Diagnosing of heart diseases using average K-nearest neighbor algorithm of data mining," in *Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 3099–3103, IEEE, New Delhi, India, March 2016.

[15] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: a review," *Egyptian Informatics Journal*, vol. 19, no. 3, pp. 179–189, 2018.

[16] I. Ibrahim and A. Abdulazeez, "The role of machine learning algorithms for diagnosing diseases," *Journal of Applied Science and Technology Trends*, vol. 2, no. 1, pp. 10–19, 2021.

[17] V. D. Soni, "Chronic disease detection model using machine learning techniques," *International Journal of Scientific & Technology Research*, vol. 9, no. 9, pp. 262–266, 2020.

[18] P. Ghosh, S. Azam, A. Karim, M. Jonkman, and M. D. Z. Hasan, "Use of efficient machine learning techniques in the identification of patients with heart diseases," in *Proceedings of the 2021 the 5th International Conference on Information System and Data Mining*, pp. 14–20, Silicon Valley USA, May 2021.

[19] Y. Chang and X. Chen, "Estimation of Chronic Illness Severity Based on Machine Learning Methods," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1999284, 13 pages, 2021.

[20] T. H. H. Aldhyani, A S Alshebami, and M. Y. Alzahrani, "Soft clustering for enhancing the diagnosis of chronic diseases over machine learning algorithms," *Journal of Healthcare Engineering*, vol. 2020, Article ID 4984967, 16 pages, 2020.

[21] H. Fröhlich, R. Balling, N. Beerenwinkel et al., "From hype to reality: data science enabling personalized medicine," *BMC Medicine*, vol. 16, no. 1, pp. 1–15, 2018.

[22] S. Ganiger and K. M. M. Rajashekharaiah, "Chronic diseases diagnosis using machine learning," in *Proceedings of the 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*, pp. 1–6, IEEE, Kottayam, India, December 2018.

[23] D. J. Park, M. W. Park, H. Lee, Y.-J. Kim, Y. Kim, and Y. H. Park, "Development of machine learning model for diagnostic disease prediction based on laboratory tests," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.

[24] J. Mishra and S. Tarar, "Chronic disease prediction using deep learning," in *Proceedings of the International Conference on Advances in Computing and Data Sciences*, pp. 201–11, Springer, Valletta, Malta, April 2020.

[25] E. Jeong, S. Osmundson, C. Gao, D. R. V. Edwards, M. Bradley, and Y. Chen, "Learning the impact of acute and chronic diseases on forecasting neonatal encephalopathy," *Computer Methods and Programs in Biomedicine*, vol. 211, Article ID 106397, 2021.

[26] F. Ceccarelli, M. Sciandrone, C. Perricone et al., "Prediction of chronic damage in systemic lupus erythematosus by using machine-learning models," *PLoS One*, vol. 12, no. 3, Article ID e0174200, 2017.

[27] Md M. Mottalib, J. C. Jones-Smith, B. Sheridan, and R. Beheshti, "Identifying the risks of chronic diseases using BMI trajectories," 2021, https://arxiv.org/abs/2111.05385.

[28] S. Agarwal, C. Prabha, and M. Gupta, "Chronic diseases prediction using machine learning–A review," *Annals of the Romanian Society for Cell Biology*, vol. 25, pp. 3495–3511, 2021.

[29] T. Takura, K. H. Goto, and A. Honda, "Development of a predictive model for integrated medical and long-term care resource consumption based on health behaviour: application of healthcare big data of patients with circulatory diseases," *BMC Medicine*, vol. 19, no. 1, pp. 1–16, 2021.

[30] D. Zufferey, T. Hofer, H. Jean, M. Schumacher, R. Ingold, and S. Bromuri, "Performance comparison of multi-label learning algorithms on clinical data for chronic diseases," *Computers in Biology and Medicine*, vol. 65, pp. 34–43, 2015.

[31] L. Beretta and A. Santaniello, "Nearest neighbor imputation algorithms: a critical evaluation," *BMC Medical Informatics and Decision Making*, vol. 16, no. 3, pp. 197–208, 2016.

[32] S.-K. Lee, Y.-J. Son, J. Kim et al., "Prediction model for health-related quality of life of elderly with chronic diseases using machine learning techniques," *Healthcare Informatics Research*, vol. 20, no. 2, pp. 125–134, 2014.

[33] K. Deepika and S. Seema, "Predictive analytics to prevent and control chronic diseases," in *Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (ICATccT)*, pp. 381–86, IEEE, Bangalore, India, July 2016.

[34] J. Peng, E. C. Jury, P. Donnes, and C. Ciurtin, "Machine learning techniques for personalised medicine approaches in immune-mediated chronic inflammatory diseases: applications and challenges," *Frontiers in Pharmacology*, vol. 12, p. 2667, 2021.

[35] S. N. Induja and C. G. Raji, "Computational methods for predicting chronic disease in healthcare communities," in *Proceedings of the 2019 International Conference on Data*

Science and Communication (IconDSC), pp. 1–6, IEEE, Bangalore, India, March 2019.

[36] A. Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit, S. Suwannawach, and N. Ninchawee, "Predictive analytics for chronic kidney disease using machine learning techniques," in *Proceedings of the 2016 Management and Innovation Technology International Conference (MITicon)*, October 2016.

[37] B. Gudeti, S. Mishra, S. Malik, T. F Fernandez, A. K. Tyagi, and S. Kumari, "A novel approach to predict chronic kidney disease using machine learning algorithms," in *Proceedings of the 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1630–1635, IEEE, Coimbatore, India, November 2020.

[38] K. Arumugam, M. Naved, P. P. Shinde, O. Leiva-Chauca, A. Huaman-Osorio, and T. Gonzales-Yanac, "Multiple disease prediction using machine learning algorithms," *Materials Today Proceedings*, 2021.

[39] K. Ng, A. Ghoting, S. R. Steinhubl, W. F. Stewart, M. Bradley, and J. Sun, "PARAMO: a PARAllel predictive MOdeling platform for healthcare analytic research using electronic health records," *Journal of Biomedical Informatics*, vol. 48, pp. 160–170, 2014.

[40] A. Choudhury, E. Renjilian, and O. Asan, "Use of machine learning in geriatric clinical care for chronic diseases: a systematic literature review," *JAMIA Open*, vol. 3, no. 3, pp. 459–471, 2020.

[41] A. S. Hussein, W. M. Omar, Li Xue, and M. Ati, "Efficient chronic disease diagnosis prediction and recommendation system," in *Proceedings of the 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences*, pp. 209–14, IEEE, Langkawi, Malaysia, Decemer 2012.

[42] J.-H. Lin and Y.-J. Hu, "Application of machine learning to immune disease prediction," *International Journal of Engineering and Innovative Technology*, vol. 7, no. 11, pp. 38–42, 2018.

[43] M. Khalilia, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," *BMC Medical Informatics and Decision Making*, vol. 11, no. 1, pp. 1–13, 2011.

[44] S. Hegde and M. R. Mundada, "Early prediction of chronic disease using an efficient machine learning algorithm through adaptive probabilistic divergence based feature selection approach," *International Journal of Pervasive Computing and Communications*, vol. 17, pp. 20–36, 2020.

[45] S. Programming, "Retracted:: enhancing health risk prediction with deep learning on big data and revised fusion node paradigm," *Hindawi*, vol. 2019, Article ID 9757658, 1 page, 2019.

[46] S. Vinitha, S. Sweetlin, H. Vinusha, and S. Sajini, "Disease prediction using machine learning over big data," *Computer Science and Engineering: International Journal*, vol. 8, no. 1, pp. 1–8, 2018.

[47] S. Y. Yashfi, Md A. Islam, N. Sakib et al., "Risk prediction of chronic kidney disease using machine learning algorithms," in *Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–5, IEEE, Kharagpur, India, July 2020.

[48] A. I. Taloba, A. A. Sewisy, and Y. A. Dawood, "Accuracy enhancement scaling factor of viola- jones using genetic algorithms," in *Proceedings of the 14th International Computer Engineering Conference (ICENCO)*, pp. 209–212, Cairo, Egypt, Decembe 2018.

[49] R. Al-Mashhadani, Gamal Alkawsi, Y. Baashar et al., "Deep learning methods for solar fault detection and classification: a review," *Information Sciences Letters*, vol. 10, no. 2, pp. 323–331, 2021.

[50] A. Al-Sammarraee and N. Alshareeda, "The role of artificial intelligence by using automatic accounting information system in supporting the quality of financial statement," *Information Sciences Letters*, vol. 10, no. 2, pp. 223–254, 2021.

[51] A. Ismail, M. L. Salem, A. Elkholy, W. Elmashad, A. Gomaa, and M. Ali, "In-silico analysis of protein receptors contributing to SARS- COV-2 high infectivity," *Information Sciences Letters*, vol. 10, no. 3, pp. 561–570, 2021.

[52] O. Y. M. Al-Rawi, "Wisam subhi Al-dayyeni, and ibrahim reda, "COVID-19 impact on education and work in the kingdom of Bahrain: survey study"," *Information Sciences Letters*, vol. 10, no. No. 3, pp. 427–433, 2021.

[53] Md R. Hoque and M. Sajedur Rahman, "Predictive modelling for chronic disease: machine learning approach," in *Proceedings of the 2020 the 4th International Conference on Compute and Data Analysis*, pp. 97–101, Silicon Valley, CA, USA, March 2020.

[54] Md E. Hossain, *Predictive Modelling of the Comorbidity of Chronic Diseases: A Network and Machine Learning Approach*, PhD Thesis, University of Sydney, Camperdown NSW, Australia, 2020.

[55] X. Zhang, H. Zhao, S. Zhang, and R. Li, "A novel deep neural network model for multi-label chronic disease prediction," *Frontiers in Genetics*, vol. 10, p. 351, 2019.

[56] R. Avanzato and F. Beritelli, "Automatic ECG diagnosis using convolutional neural network," *Electronics*, vol. 9, no. 6, p. 951, 2020.

[57] G. Kaur and A. Sharma, "Predict chronic kidney disease using data mining algorithms in hadoop," in *Proceedings of the 2017 International Conference on Inventive Computing and Informatics (ICICI)*, pp. 973–979, IEEE, Coimbatore, India, November 2017.