

RESEARCH ARTICLE

A Novel Artificial Spider Monkey Based Random Forest Hybrid Framework for Monitoring and Predictive Diagnoses of Patients Healthcare

REYAZUR RASHID IRSHAD¹, SHAHID HUSSAIN^{1,2}, IHTISHAM HUSSAIN³, AHMED ABDO ALATTAB¹, ADIL YOUSIF¹, OMAR ALI SALEH ALSAIARI¹, AND ELSHAREEF IBRAHIM IDREES IBRAHIM⁴

¹Department of Computer Science, College of Science and Arts Sharurah, Najran University, Najran 68341, Saudi Arabia

²Innovation Value Institute (IVI), School of Business, National University of Ireland Maynooth (NUIM), Maynooth, Co. Kildare, W23 F2H6 Ireland

³Department of Computer Science, Abdul Wali Khan University of Mardan (AWKUM), Shaheed Rashid Hussain (SRH) Campus, Nowshera, Pabbi 24210, Pakistan

⁴Department of Mathematics, College of Science and Arts Sharurah, Najran University, Najran 68341, Saudi Arabia

Corresponding author: Shahid Hussain (shahid.hussain@mu.ie)

The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Research Priorities and Najran Research funding program. Grant Code (NU/NRP/SERC/12/11).

ABSTRACT Early diagnosis of diseases such as cancer, cardiovascular, diabetes, HIV, AIDS, Lyme, and tuberculosis can enable timely treatment, which enhances efficacy and helps reduce the severity and mortality while lowering overall healthcare costs. Predictive analytics is an emerging approach that comprehends trends in patient data to identify patterns in patient data, detect subtle signs of diseases, and enable early diagnosis for more effective treatment. However, traditional predictive diagnosis approaches rely on statistical models that are contingent on the data's average pattern and are unable to capture the data's inherent patterns, leading to ineffective and inaccurate diagnoses. In this paper, we introduce a novel Artificial Spider Monkey-based Random Forest (ASM-RF) hybrid framework that combines the predictive analytics of a Random Forest algorithm with Artificial Intelligence to evaluate patient health data, spot patterns, diagnose modest indications, and automate intelligent decisions for enhancing the healthcare system. The proposed ASM-RF hybrid framework employs a fitness function to evaluate the spider monkey's performance at the classification layer and update accuracy and recall, resulting in more accurate patient disease diagnoses and automating timely treatment decisions for improving the overall healthcare system. Moreover, we use Identity-based Encryption (IBE), which enables the encryption of data with private and public keys coupled with users' identities as the encryption keys, assisting in enhancing the security of the healthcare system. A dataset collected from three different IoT sensor devices, ten participants, and twelve activities is employed to simulate the proposed ASM-RF hybrid framework, which is then contrasted to cutting-edge predictive diagnostic algorithms in terms of accuracy, precision, Area under the Curve, execution time, recall, and F-measure. The proposed method exhibits superior performance when compared to conventional methods, as evidenced by its exceptional accuracy (99.52%), precision (99.12%), Area Under the Curve (AUC) (99.00%), recall (99.22%), F-measure (97.12%), and significantly reduced execution time (6s).

INDEX TERMS Health monitoring, hybrid IoT framework, identity-based encryption, meta-heuristic, predictive analytics, random forest, spider monkey optimization.

I. INTRODUCTION

The associate editor coordinating the review of this manuscript and approving it for publication was Alessandro Floris¹.

Early disease diagnosis is a critical part of preventive medicine since it enables the detection and treatment of

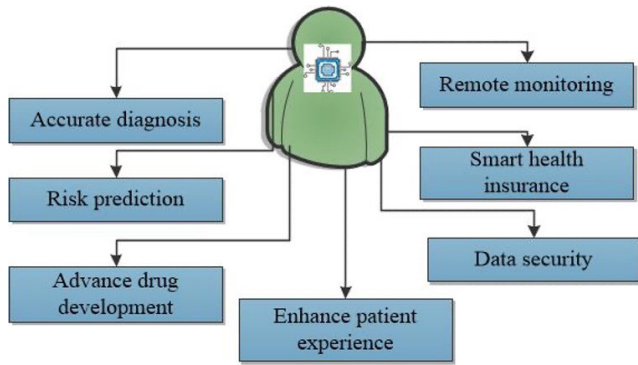


FIGURE 1. An illustration of the implication and benefits of artificial intelligence in healthcare.

medical conditions at the earliest possible stage and aids healthcare professionals in keeping up with the most recent methods and treatments for a range of ailments [1]. Meanwhile, since many diseases do not exhibit symptoms until they have advanced further, early detection is one of the main challenges in spotting the disease before it has progressed too far [2]. The global healthcare system is experiencing enormous challenges, such as high expenses, limited access, an aging population, and a lack of safety equipment, all of which have an impact on the healthcare system from both a patient's healthcare and a financial standpoint [3]. The healthcare industry is developing, as indicated by rising healthcare expenditures in general and a growing lack of healthcare personnel [4], [5], [6]. The expansion of sensing technologies and wearable medical devices, as well as communication protocols, has enabled healthcare providers to improve their ability to monitor patients in real-time while managing their records, which include treatment, history, and other lab results [7]. Besides, the advancement of technology and the incorporation of artificial intelligence [8], [9] into the healthcare system offers the potential to address a wide range of issues, including early diagnosis, real-time monitoring, and financial burden reduction [10], [11]. Monitoring sensors and medical wearable devices generate huge amounts of data [12], which provides opportunities to analyze the data and anticipate diseases for early decision-making purposes, where artificial intelligence-based machine learning (ML) frameworks [13], [14] play a significant role, as shown in Figure 1.

By utilizing Internet-of-Things (IoT) sensors [15], [16] and wearable medical devices to monitor and collect patients' data in real-time, AI-based algorithms can provide high-precision, speed, flexibility, and timeliness when it comes to dealing with healthcare systems [17]. This implies that as fog computing, IoT, cloud computing, and ML grow increasingly capable of connecting to individuals' environments and lifestyles, the possibility of diagnosing various physiologic ailments at an early stage is increasing [18], [19]. Real-time monitoring of a patient's health and the early diagnosis of derangements may allow for speedy and effective treatment,

lower patient care costs, and improve patient outcomes all while lowering the cost of the healthcare system [20], [21]. Thus, AI-based ML algorithms have the potential to minimize high costs, lengthy development times, and disease detection at an early stage [22], but security, data privacy, and accurate identification remain major challenges in controlling the healthcare system through technology. Several strategies have been developed to improve the functionality of healthcare systems in detecting and diagnosis of the various diseases [23], [24], [25], [26], and the security algorithms [27], [28], [29] to safeguard the sensitive healthcare data, but most of these techniques rely on traditional methods that depend on statistical models for analyzing data and evaluating average cases, resulting in inadequate disease diagnosis. The current prevailing paradigm heavily relies on the presence of symptoms and interpersonal contact. Previous studies have utilized various optimization algorithms such as Ant Lion Optimization (ALO) [30], Biography Based Optimization (BBO) [31], Particle Swarm Optimization (PSO), and Grey Wolf Optimization. However, these earlier optimization methods exhibit drawbacks including low accuracy, limited local searching capabilities, and slow convergence rates. In other previous research, the Gaussian Mutation–Spider Monkey Optimization (GM-SMO) model was developed for remote sensing scenario classification [32]. Additionally, Spider Monkey Optimization (SMO) has been combined with deep learning methods for intrusion detection [33]. Given the needs of the healthcare system and the benefits of AI-based ML algorithms, it is necessary to develop an AI-enabled healthcare framework that supports real-time monitoring, accurate early disease identification, cost reduction, and high levels of security so that healthcare organizations may have trust in the technology. To the best of the author's knowledge, the ASM-RF hybrid framework and the implementation of Identity-based Encryption for disease prediction and securing sensitive healthcare data have not been previously presented. In this work, we concentrated on the development of an AI-enabled healthcare system with the inclusion of security mechanisms to provide reliable and cost-effective monitoring and diagnosis. The proposed platform aims to eliminate diagnostic testing processes, hospital stays, and doctor office visits by automating healthcare management using AI-enabled ML algorithms. As the ASM algorithm introduces intelligence and adaptability to the feature selection process, utilizing a meta-heuristic approach inspired by spider monkey behavior to prioritize and select important features, resulting in improved decision tree construction, reduced inclusion of irrelevant features, and enhanced overall performance and accuracy of the ensemble learning process [34]. The suggested technique holds significant promise in the healthcare domain by empowering clinicians and healthcare professionals to achieve early detection and timely treatment of critical health issues. Moreover, the incorporation of Identity-Based Encryption ensures the robust security of healthcare data, safeguarding it against unauthorized

access and potential breaches. The following summarizes the key contribution of this paper.

- We present a novel Artificial Spider ASM-RF hybrid framework that integrates the predictive analytics of a Random Forest algorithm with Artificial Intelligence to analyze patient health data, find trends, recognize subtle symptoms, and automate smart decisions for enhancing the healthcare system. The developed ASM-RF hybrid framework permits the fitness function to evaluate the spider monkey's performance at the classification layer and update the disease detection, enabling the automation of prompt treatment decisions and supporting the lowering of the workload of physicians and expense of the healthcare system. Moreover, we established a synergistic coupling between the ASM and the RF algorithms, where the ASM algorithm intelligently selects optimal features to enhance the performance and accuracy of the RF algorithm in disease prediction and diagnosis.
- Given the sensitivity and complexity of the healthcare system, we utilize Identity-based Encryption (IBE), a cryptographic technique that allows the encryption of data by combining private and public keys with users' identities as encryption keys. With IBE, each user is assigned a unique identity, and a securely stored private key generated based on this identity is used for decrypting data encrypted with the corresponding public key. This approach not only ensures the confidentiality and integrity of healthcare data but also simplifies the key management process, as there is no need to distribute and manage individual encryption keys for each user. By leveraging IBE, our system strengthens the security measures in the healthcare system, protecting sensitive data from unauthorized access and ensuring the privacy of patients' information.
- We simulated the proposed ASM-RF hybrid framework using a dataset obtained from three different IoT sensor devices, ten participants, and twelve activities, and the results are compared to the cutting-edge Artificial Intelligence and the Internet of Things (AI-IoT), Secure IoT-based Smart Healthcare Monitoring System (SIoT-SHMS), Controlling Safety of AI-based Healthcare Systems (CS-AI-HS), Health Monitoring System utilizing Intelligent Task Mapping (HMS-ITM), and Fog Computing-based Intelligent Healthcare System (FC-HIS) predictive diagnostic algorithms in terms of accuracy, precision, Area under the Curve, execution time, recall, and F-measure.

The following is an outline of the contents of this paper: Section II describes the paper's literature review. The investigated problem is addressed in more detail in Section III. Section IV outlines the proposed hybrid approach. Section V will go over the acquired data and do a comparative analysis. The summary appears in Section VI.

II. LITERATURE SURVEY

Amit Kishor and Chinmay [35] proposed an ML-based healthcare framework for effectively predicting the early and late outcomes of certain diseases. This paper examines the use of Artificial Intelligence and the Internet of Things to create a Healthcare 4.0 monitoring system. It explains how this new system can enable better access to healthcare services and reduce healthcare expenditure. Additionally, it discusses the potential challenges that could be faced in implementing this system. This system predicts nine severe diseases utilizing seven ML classification techniques, enabling clinicians to make early diagnoses more effectively. However, it is important to note that this paper primarily focuses on disease prediction and does not extensively cover categorization concerns and potential security risks associated with the collection and storage of patient data.

Hesham and Mustafa [36] proposed an integrative AI strategy based on fuzzy systems and neural networks to safeguard Healthcare Management Systems (HMS). This article covers the use of secure IoT communications for Smart Healthcare Monitoring Systems, with an emphasis on the development of an IoT architecture designed to collect and exchange patient data in a secure manner. Furthermore, it investigates the system's benefits, such as improved healthcare services and lower healthcare costs. One shortcoming of this paper is that it does not address potential legal and privacy issues associated with the usage of IoT in the healthcare system.

Ghazal [37] developed an IoT-based AL technique for healthcare security, as well as a Wireless Sensor Network (WSN) that connects the physical and virtual worlds using IoT approaches. This article investigates the application of the Internet of Things and Artificial Intelligence for healthcare security, using AI-based capabilities for securely gathering and maintaining patient data. It goes on to highlight the possible benefits of this system, such as improved healthcare services and better patient outcomes. Therefore, the created framework could keep track of, secure, and encrypt patient data kept on the cloud. This paper, on the other hand, did not provide enough information about security methods for keeping and protecting patient data due to extended encryption and decryption periods.

Reza et al. [38] devised a safety control system to reduce the risk of medication mistakes. This article investigates how to control the safety of AI-based systems in healthcare by reviewing the existing regulatory framework and standards to identify potential dangers connected with the use of AI in health care and presents solutions for minimizing these risks. However, it solely addresses the safety of AI-based systems in healthcare and ignores other critical issues like patient privacy and data security.

Iqbal et al. [39] in another work proposed an intelligent task-mapping mechanism to improve health monitoring systems for older patients in a closed-loop healthcare context. It implies that the application of intelligent task mapping can deliver a more affordable method of monitoring older

people and allow them to receive individualized treatment. A further suggestion made in the paper is that using intelligent task mapping could help medical professionals understand the health status of their senior patients and help them decide how best to care for them. The fundamental shortcoming of the paper is that it lacks detail on the potential obstacles and costs associated with implementing an intelligent job mapping mechanism in a closed-loop healthcare context.

Halder and Datta [40] proposed an artificial intelligence-based approach for detecting COVID-19 in CT-scan images. The research discusses how the model was trained using a transfer learning strategy and has shown promising results in detecting COVID-19 infections from lung CT-scan pictures. The paper goes on to explore the suggested model's possible application in clinical settings as well as its future potential for boosting COVID-19 diagnosis accuracy. The paper's primary flaw is that it does not assess the model's efficacy in diagnosing other conditions, such as other varieties of respiratory infections.

Bui et al. [41] describe a novel corotational cut finite element (CCFE) approach for real-time surgical simulation. The suggested CCFE approach was used to mimic needle insertion with realistic tissue deformations and yielded promising results in terms of accuracy and computational efficiency when compared to existing methods. The research also explored how this new technique can be used to improve existing medical training tools and increase the accuracy of medical simulations. The paper's key shortcoming is that it only addresses needle insertion simulation and no other sorts of surgical simulations.

Zhang et al. [42] suggested a lightweight convolutional neural network (CNN) based on a transfer learning technique for recognizing COVID-19. When compared to existing approaches, the proposed model produced promising results in terms of accuracy and computing complexity. The paper discusses the potential of the model to improve the accuracy and efficiency of COVID-19 diagnosis in clinical settings. However, it is important to acknowledge that the paper's primary limitation is the lack of assessment regarding the model's accuracy in detecting other respiratory infectious disorders.

El-Sappagh et al. [43] proposed an intelligent healthcare system for cardiac disease prediction, employing ensemble deep learning techniques and feature combination methodologies. To optimize system performance and alleviate computational load, they applied an information gain approach to eliminate redundant features and select essential ones. Their collective deep learning model was trained to forecast cardiac disease. Nevertheless, this model exhibited a notable limitation with regards to its accuracy.

El-Sappagh et al. [44] in their other work introduced an innovative medical surveillance system, harnessing data mining methods, ontologies, and bidirectional long short-term memory (Bi-LSTM). This system leverages cloud computing infrastructure and employs a robust big data statistics engine to accurately maintain and analyze healthcare data, thereby

significantly improving classification accuracy. Furthermore, the suggested methodology encompasses the categorization of patients' health conditions based on crucial medical information, including blood pressure, diabetes indicators, psychological state, and drug reviews.

Employing the stacking approach, El-Sappagh et al. [45] presented an innovative ensemble learning system for the prognostication of Alzheimer's disease, which integrates a diverse set of base learners within a unified framework. The selection of ensemble constituents involves a comprehensive exploration of the interplay between accuracy and diversification. Notably, the proposed model exhibits exceptional performance even in the absence of neuroimaging data, thus rendering it suitable for cost-effective deployment in healthcare settings.

The literature review portrays that the majority of the existing works primarily focus on addressing specific diseases, thereby neglecting the impact on individuals affected by other diseases in remote areas. Furthermore, certain studies have only analyzed limited aspects of security concerns. It is worth noting that previous models have exhibited poor scalability and accuracy in terms of data transmission security and disease prediction. The preservation and safeguarding of patient information, coupled with security concerns related to privacy, present significant challenges in the context of digitization. Given the massive volume of data stored within hospital systems, there is a heightened risk of potential data breaches, which not only compromise the accuracy of prediction performance but also escalate system costs [38]. Therefore, there arises an imperative for the development of a novel algorithm that can ensure both precision and cost reduction in healthcare system predictions while prioritizing security. In response to this demand, the present study introduces an innovative approach that integrates the ASM optimization methodology with the RF algorithm and the BE security algorithm.

III. THE PROBLEM DEFINITION

The main concerns in the healthcare environment include high prices, limited accessibility, security, privacy, a shortage of medical professionals, and safety precautions. The security of healthcare and patient monitoring are essential responsibilities for managing the healthcare system. However, due to a large number of data, misclassification, and detection errors are common, making the process of tracking a patient's health condition take longer. To enhance the efficiency of healthcare monitoring systems, a brand-new machine learning framework based on optimization is being developed. It improves healthcare monitoring system performance and incorporates cryptographic approaches to promote healthcare data security.

IV. THE METHODOLOGY OF THE PROPOSED ARTIFICIAL SPIDER MONKEY-BASED RANDOM FOREST HYBRID FRAMEWORK

The main objective of the healthcare monitoring system is to conserve the cost of healthcare by lowering hospital

stays, diagnostic testing procedures, and doctor office visits. Nonetheless, monitoring and preserving healthcare data from unauthorized access is a challenging task. As a result, a range of patient-related data is collected through IoT devices and used to train the algorithm. The use of identity-based encryption while storing data in the cloud ensures data security. Therefore, the suggested ASM-RF system is a cutting-edge platform that can allow ongoing patient health monitoring. The approach is intended to harness the strengths of both the Artificial Spider Monkey algorithm and the Random Forest classifier, such as the resilience of the ASM model and the precision of the Random Forest classifier, to tackle the disease prediction problem more correctly and effectively as shown in Figure 2. The figure indicates that the input data was examined and filtered throughout the preparation process to remove noise, errors, and missing values. Following that, using feature extraction, relevant features are retrieved from the dataset. The classification layer's Spider Monkey (SM) fitness function is therefore updated to keep track of patient status data and classify the patient's state of health. First, numerous patient datasets are compiled from a variety of sources, including hospitals, labs, distant sites, and home data collected via IoT devices and other sensors such as pulse oximeter sensors, temperature sensors, pressure sensors, accelerometers, and heart rate sensors. Furthermore, to secure the confidentiality and integrity of sensitive patient data, identity-based encryption is used, which generates a user's public key from the user's identification parameters, such as their name, email address, etc., that ensure the security of patient-specific datasets in the cloud. Additionally, the IBE aids in authenticating patient-provider communications and preventing the abuse of personal data. Finally, the updated SM fitness in the classification layer for monitoring the patient's health condition via the SM's behavior is used to properly identify and classify the patient's health status while offering rapid processing.

A. THE DESCRIPTION OF THE DATASET

The dataset used for the proposed ASM-RF has been obtained from the Kaggle¹ website. The Kaggle Mobile Health dataset, which contains over 120 million healthcare-related records such as prescriptions, disease diagnoses, test findings, and health insurance information, is used to simulate the planned ASM-RF. The dataset covers a variety of data types, including medical information, patient demographics, billing information, and laboratory tests, which can be utilized to build predictive models for healthcare applications. The dataset was meticulously curated using three sensor devices, involving 10 participants and encompassing twelve distinct activities. These sensors were strategically positioned to monitor the acceleration, rate of rotation, and magnetic field direction of specific body parts implanted on the patient's right wrist, chest, and left ankle. The dataset comprises a comprehensive

¹<https://www.kaggle.com/datasets/kaushil268/disease-prediction-using-machine-learning>

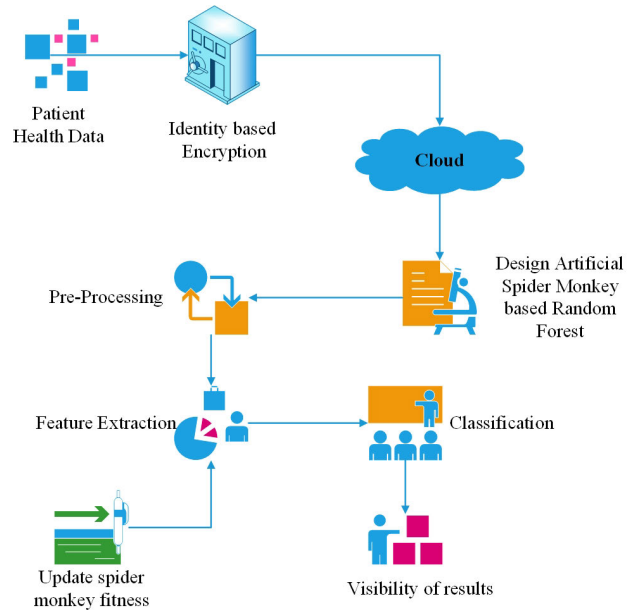


FIGURE 2. A representation of the architecture of the proposed Artificial Spider Monkey-based Random Forest Hybrid Framework, highlighted the interlink between the different components in healthcare system.

collection of 132 parameters associated with 42 diseases, ensuring a robust and diverse data representation for analysis and exploration. Furthermore, the sensors mounted on the patient's chest generate 2-lead ECG readings that can be employed for basic cardiac monitoring and the identification of various arrhythmias. The dataset is divided into 0.7 and 0.3 portions, with the first 70% being used to train the algorithm and the remaining 30% being used to test its performance. The model is trained and evaluated using 100 iterations, and the related statistics are collected for patient health status prediction.

B. THE MECHANISM OF THE IDENTITY-BASED ENCRYPTION FOR THE HEALTHCARE DATA

The IBE is a type of encryption that uses a person's identity as the encryption key. It is a successful approach for protecting sensitive healthcare data, such as patient medical records, and assures that only the intended receiver has access to the information. IBE has the benefit of enabling data encryption without the necessity for recipient public key knowledge. This makes it possible for service providers, including healthcare providers, to store and securely share sensitive data without worrying about data breaches. Additionally, it provides the option of rescinding access to data if necessary, and both implementing and maintaining the approach is not too difficult. The created architecture's primary objective is to safeguard data from attacks, and external parties also enhance cloud security performance by constantly scanning for attacks that target the healthcare industry. Several patient healthcare datasets are first changed to suit the proposed model, and the data is then protected using the IBE technique.

Using cryptography and machine learning techniques, the data is protected from attacks and unwanted access, while the created health records are adjusted to the intended model. Furthermore, the IBE method converts the original data or plain text into ciphertext, which is then transported to the cloud and stored as ciphertext, where it may only be viewed by those who have been granted permission by the data owners. The following main steps are performed during the identity-based encryption process.

1) KEY GENERATION

The key is generated using a pair of polynomials known as the private key polynomial and the public key polynomial. The private key polynomial is only known to user, whereas the public key polynomial can be computed by anybody who has access to the user's identity. IBE encrypts and decrypts data by multiplying the plaintext by the public key polynomial, whereas the decryption process requires multiplying the ciphertext by the private key polynomial. The generated private value and session key are secured by generating a random value for the patient's identity and as a result, the produced is used to determine the associated private key as presented in Eq. (1).

$$K_g = \left(m_k \left(P' \prod_{i=k}^n r_i \right), Sv_a \right) \quad (1)$$

where r_i represents the i th patient identity, m_k , P' , and Sv_a , are the primary, public, and session keys, respectively.

2) DATA ENCRYPTION

The data encryption is performed through the implication of the primary key m_k and the patient identity information r_i that converts plain text into ciphertext. The ciphertext is denoted by $k(t)$ the plain text is represented by $v(t)$ and the public key is denoted by P_k . The transformation process of plan text to the ciphertext is performed through Eq. (2).

$$encrypt_{data} = \left(v(t)(m_k)P_k \left(P' \prod_{i=k}^n r_i \right), k(t) \right) \quad (2)$$

The encryption is reliant on the validity of the public key P_k such that if the key is valid, it transforms the plain text to ciphertext otherwise the encryption will fail. The data is thus safeguarded from potential assaults and unauthorized access by the private and public key combination and is kept in the cloud system.

3) DATA DECRYPTION

The data at the receiving end is decrypted by multiplying the data ciphertext by the private key polynomial d_k , which is known only to the user, to yield the plaintext, as shown in Eq. (3).

$$decrypt_{data} = \left(d_k \cdot M_k, \frac{k(t)(m_k(1,2,\dots,n)) \left(P' \prod_{i=k}^n r_i \right)}{k(t)(m_k) \left(P' \prod_{i=k}^n r_i \right)}, v(t) \right) \quad (3)$$

where $M_k = m_k, m_{k1}, m_{k2}, m_{k3}$ are the primary keys, d_k is the private key polynomial, $k(t)$ is the cipher text, P' is the public key, and r_i is the i th identities of patients. When sending the data to the cloud system for storage and additional processing, the security of the encrypted data is preserved using the private key.

C. THE MECHANISM OF THE ARTIFICIAL SPIDER MONKEY-BASED RANDOM FOREST

The ASM-RF model leverages a heuristic artificial spider monkey search that employs evolutionary algorithms to determine features that best identify trends in the patients' data, ultimately enhancing the system's classification results for healthcare monitoring. The input layer gets multiple patient data points, and the health condition of each patient is then classified based on their blood pressure, blood sugar, and other dehydration parameters, with the output layer reflecting the patient's current health status.

It maintains the spider monkey fitness function by adoptively updating it during the classification phase. Spider monkeys employ postures and attitudes, such as sexually receptive and aggressive postures, to communicate their intents and observations with one another while traversing long distances using a particular call that resembles a horse's whinny. Spider monkeys can socialize, avoid danger, share food, and converse using this long-distance communication. They primarily employ visual and oral communication to engage with other group members, and spider monkey foraging activity is divided into four stages in the established approach. The group begins food collection initially and measures its closeness to the food, while the second step involves group members updating their locations and re-evaluating the distances from the food sources to calculate the distance from the foods. In the third phase, the local leader updates the group's best position, and if that location is not updated after a specified number of times, every member of that group starts looking for food in different directions. In the fourth phase, the global leader maintains its ever-best position, and in the event of stagnation, it divides the membership into smaller size subgroups. As a consequence, the developed system continually monitors behavioral and physical changes while also categorizing each patient's health condition using foraging behavior.

Furthermore, random forest is a meta-estimator that combines numerous decision tree classifiers to improve prediction accuracy while minimizing overfitting. It generated decision trees from numerous samples, and the classification and regression results of the decision trees were utilized to calculate the majority vote.

1) DATA PREPROCESSING

The obtained patient datasets are in raw form and cannot be adequately suited to the algorithm; consequently, it must be evaluated and filtered to remove training errors, mistakes, and missing values. Data preprocessing is a set of processes used to transform raw data into clean, machine-readable data.

Typically, it entails cleaning, smoothing, encoding, and/or altering data to make it more amenable for analysis or other operations. In this study, the data cleaning and preprocessing are carried out using Eq. (4).

$$P(r) = \frac{i(t)r_i - n(o)}{r_i} \quad (4)$$

where, $n(o)$ represents the noise and error present in the received data, $i(t)$ is characterized as received data, and r_i denotes as patient identity (ID).

2) FEATURE EXTRACTION

Given the diverse range of data types, features, and record counts within the database, the task of extracting meaningful features from preprocessed data is commonly referred to as feature extraction. This process typically involves selecting and modifying variables that effectively capture the most relevant information from the data while eliminating redundant or unnecessary features. In this work, we have adapted the 10-fold cross-validation strategy for the feature selection, which eliminates the overfitting issues. The 10-fold cross-validation splits the dataset into ten subsets, enabling testing and training on different sets of data. In each iteration, one subset is used for testing while the remaining subsets are considered for training. This approach minimizes the input features from a pool of 132 features by selecting the most optimal ones, including age, gender, symptoms, blood pressure, time, blood glucose level, activity, heart rate, body temperature, and weight scale. We utilized internal cross-validation, a widely employed technique for feature selection, with the objective of identifying the most influential features that substantially contribute to the predictive capacity of a model [46]. To execute this process, the data is initially partitioned into training and validation sets, ensuring a representative distribution of samples. Subsequently, for each distinct feature subset, the model is trained exclusively on the training set, and its performance is then evaluated using the validation set. To minimize bias, this training and evaluation procedure is repeated multiple times, employing different random splits of the data. Finally, based on the chosen accuracy evaluation the feature subset that exhibits the most favorable performance is selected as the optimal choice, as shown in Figure 3. This meticulous approach enables the identification of a feature subset that demonstrates robust generalization capabilities and enhances the overall predictive power of the model. By incorporating these selected features, the accuracy of machine learning models can be significantly enhanced. The mathematical representation of the appropriate feature selection is presented in Eq. (5).

$$E(f) = \sum_{i=1}^n i_f \times P_f(i) \times R_f(n) \frac{r_i}{P_r} \quad (5)$$

where, P_f denotes the extracted relevant features, i_f represents unimportant features, and R_f denotes feature inputs during training.

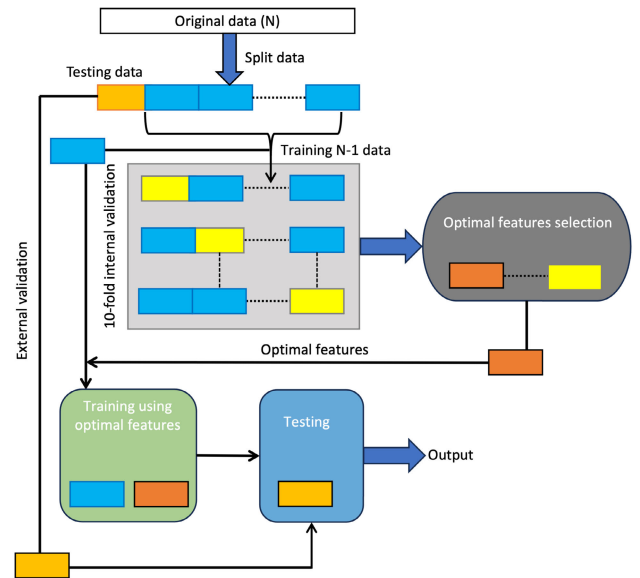


FIGURE 3. A representation of the cross-fold validation mechanism for optimal feature selection, emphasizing both internal and external validations.

3) CLASSIFICATION

The classification procedure divides the learned data into distinct matched classes by recognizing patterns and trends in the data using supervised learning. The implications of the spider monkey’s fitness function at the classification layer encourage real-time tracking of each patient’s health state as a means of enhancing classification. The spider monkey’s fitness is updated through foraging habits, and the output layer displays the patient’s overall health state. The proposed model is therefore better at predicting the health status of different patients. The categorization is carried out using Eq. (6).

$$C_l = \frac{\sum_{m=1}^n S_m(f)E(f)\left(\frac{1}{b} - b_m^2(m-1)\right)}{R_f \sum_{m=1}^n c(n)\left(\frac{1}{v} - v_m^2(m-1)\right)} \quad (6)$$

where $S_m(f)$ is the spider monkey fitness function and $c(n)$ represents the classification results of each patient. Additionally, v and v_m denote the continuous tracking of each patient’s training data. Following that, Eq. (7) is employed to establish the visibility of the patient’s health status.

$$v_i = \begin{cases} c(n) \geq S_m(f) & rand_{m,n} \\ c(n) < S_m(f) & otherwise \end{cases} \quad (7)$$

The patient’s health state is anticipated and recognized based on the numerous retrieved features from the patient’s data. Consequently, the established framework produces improved results for identifying and administering the proper drug to a patient, as well as improving performance for monitoring patient health care.

D. PSEUDOCODE OF THE PROPOSED ASM-RF-ED ALGORITHM

The proposed ASM-RF-ED is a hybrid algorithm that combines the strengths of ASM and Random Forest for disease detection, while ASM is an optimization algorithm that can be used for feature subset selection.

The first algorithm (i.e., algorithm 1), ASM-RF-ED, is a procedure for the early detection of diseases using a combination of the Artificial Spider Monkey algorithm and Random Forest. It takes in a dataset of instances with features and labels for disease status and outputs a Random Forest classifier for disease detection. The algorithm works by creating multiple decision trees using subsets of features selected by the Artificial Spider Monkey algorithm. The Random Forest classifier is constructed by aggregating these decision trees, and the overall performance of the classifier is evaluated using out-of-bag testing.

The second algorithm (i.e., algorithm 2), Artificial Spider Monkey-ASM, is a metaheuristic optimization algorithm based on the social behavior of spider monkeys. It takes the defined range of RF algorithm hyperparameters and tuned by the fitness function of the ASM algorithm optimally. The parameters used for tuning is size of trees in the forest, maximum characteristics are taken into account when splitting a node, the highest points in each decision tree, and other parameters utilized in RF for tuning purposes. Minimum number of data points in a node before it splits, minimum number of data points permitted in a leaf node, and sampling technique for data points. The algorithm works by initializing a population of spider monkeys, each representing a possible solution and updating their positions in the search space based on their fitness values. The procedure of ASM is explained as follows: First, the population of each parameters from RF is initialized. In the ASM algorithm, the current location of the ASM is modified by incorporating the knowledge and experiences of the regional leader and other individuals within the local group. The ASM updates its location only if the new position offers a higher fitness value compared to its previous location. The Eq. (8) represents the update mechanism for the position of the a^{th} ASM within the L^{th} local group.

$$S_{newab} = S_{ab} + X \times (L1_a - S_{ab}) + Y \times (S_{rb} - S_{ab}) \quad (8)$$

where $X = A(0, 1)$ and $Y = A(-1, 1)$. Following the Local Leader Phase (LLP), the global leader phase (GLP) begins. The global leader's expertise and the experiences of local group members are used to update the location of every ASM using Eq. (9).

$$S_{newab} = S_{ab} + X \times (G1_a - S_{ab}) + Y \times (S_{rb} - S_{ab}) \quad (9)$$

This updates the S_b location based on the probability value. The more opportunities there are for improvement, the better the site applicants' chances are. The global leader position is updated using a greedy selection strategy. The position of the ASM with the highest population fitness value is updated for the global leader. The worldwide leader is given the ideal

location. If no new updates are found, the global limit count is increased by one. The local group uses the greedy selection approach to update the location of the local leader. The ASM position with the best fitness in a certain local group is updated as the location of the local leader. The local leader is given the best site. If no new updates are found, the local limit count is increased by one. When a local leader fails to update its location within a predetermined local leader limit, the others in that local group either change their positions at random as per step 1 or using historical data from both the local and global leaders, using the probability as given in Eq. (10).

$$S_{newab} = S_{ab} + X \times (G1_a - S_{ab}) + Y \times (S_{rb} - Lab) \quad (10)$$

The candidate solutions are iteratively updated, and the algorithm terminates either when a termination criterion is met or when the optimal solution is found.

Algorithm 1 Artificial Spider Monkey-Based Random Forest for Early Detection of Diseases

Require: D : dataset of n instances with m features, including labels for disease status

Ensure: RF : Random Forest classifier for disease detection

```

1: procedure ASM-RF-EDD
2:  $RF \leftarrow$  empty set
3: for  $i \leftarrow 1$  to  $T$  do
4:    $D' \leftarrow$  bootstrapped sample of  $D$ 
5:    $T_i \leftarrow$  empty set
6:   for  $j \leftarrow 1$  to  $k$  do
7:      $X_j \leftarrow$  ASM( $D'$ ) ▷ Call algorithm 2
8:      $T_{ij} \leftarrow$  decision tree using  $X_j$ 
9:     add  $T_{ij}$  to  $T_i$ 
10:  end for
11:  add  $T_i$  to  $RF$ 
12: end for
13: Print results

```

The main steps of these algorithms are as follows.

- 1) Declare a procedure ASM-RF-ED that takes a dataset D as input and outputs a Random Forest classifier RF for disease detection.
- 2) Initialize an empty set RF to hold the Random Forest classifier. Loop T times to create T number of trees in the Random Forest and create a bootstrapped sample D' of the dataset D .
- 3) Initialize an empty set T_i to hold the decision trees and loop k times to create k number of decision trees in the Random Forest and call the algorithm 2 by passing the created bootstrapped sample D' .
- 4) The algorithm takes a dataset D' with instances and features as input and aims to select the best feature subset for disease detection. It initializes a population of spider monkeys and a variable X_b to keep track of the best solution so far. The algorithm iteratively calculates the fitness of each spider monkey based on disease

Algorithm 2 Artificial Spider Monkey-ASD(D)

Require: Get the dataset D with instances and features

Ensure: X_b : best solution for feature subset selection

```

1: procedure ASMD
2: Initialize population of spider monkeys  $P$ 
3: Initialize best solution  $X_b$ 
4: while termination criteria not met do
5:   for each spider monkey  $S \in P$  do
6:     calculate fitness  $f(S)$  based on disease status
7:     if  $f(S) < f(X_b)$  then
8:       update  $X_b$ 
9:     end if
10:  end for
11:  for each spider monkey  $S \in P$  do
12:    move  $S$  towards  $X_b$  based on spider monkey
13:  end for
14:  update  $P$ 
15: end while
16: return  $X_b$ 

```

status, updates X_b if a better solution is found, and moves the spider monkeys towards X_b . This iterative process continues until a termination criterion is met. Finally, the algorithm returns X_b , which represents the best feature subset selected by the spider monkeys.

- 5) Use the selected features X_j to create a decision tree T_{ij} and add the decision tree T_{ij} to the set T_i and add the set T_i to the Random Forest RF . Finally, print out the desired result.

V. RESULTS AND DISCUSSION

The proposed ASM-RF model is implemented using MATLAB and effectively employed to predict a diverse range of 42 diseases, including itching, heart disease, depression, cough, and polyuria, among others. Furthermore, to ensure the security of medical information, the IBE method is employed. The performance evaluation encompasses various metrics, such as execution time, accuracy, area under the curve (AUC), precision, F-measure, and recall, providing comprehensive insights into the model's effectiveness.

A. PERFORMANCE METRICS

The simulation results are tracked in terms of F-measure, precision, execution time, accuracy, recall, and area-under-curve performance matrices and compared to the existing algorithms including Artificial Intelligence and the Internet of Things (AI-IoT) [35], Secure IoT-based Smart Healthcare Monitoring System (SIoT-SHMS) [36], Controlling Safety of AI-based Healthcare Systems (CS-AI-HS) [38], Health Monitoring System using Intelligent Task Mapping (HMS-ITM) [39], and Fog Computing-based Intelligent Healthcare System (FC-HIS) [47] respectively.

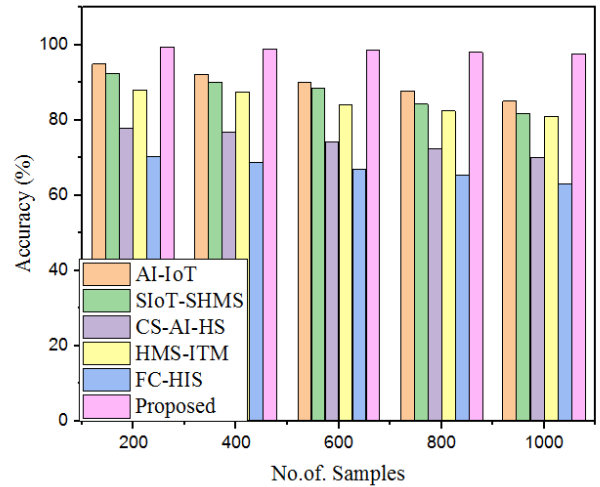


FIGURE 4. A comparison of the accuracy (measured in percentage) of the proposed ASM-RF against AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS with a varying number of sample data points.

1) ACCURACY

The accuracy parameter shows the degree of agreement between the actual and predicted values, and it is one of the most essential metrics for determining how well the model performs on unseen data. Accuracy is computed by dividing the number of correct predictions by the total number of predictions made, as specified in Eq. (11), and is commonly expressed as a percentage.

$$ac = \frac{\vec{t}p + \vec{t}n}{\vec{t}p + \vec{t}n + \vec{f}p + \vec{f}n} \tag{11}$$

where $\vec{t}p$ is the accurate analysis of the patient's health, $\vec{t}n$ is the accurate analysis of incorrect patient health status, $\vec{f}p$ is the inaccurate analysis of patient health status, and $\vec{f}n$ is the inaccurate analysis of incorrect patient health status. A comparison of the accuracy of the proposed ASM-RF against the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS using a varying number of data samples is illustrated in Figure 4. It is evident that the accuracy of the algorithms has an inverse relationship with the number of samples such as the number of data samples increases the accuracy decreases accordingly for each of the algorithms. Considering the worst case of 200 data samples, the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, FC-HIS, and the proposed ASM-RF achieve 94.90%, 92.50%, 78.00%, 88.00%, 70.34%, and 99.52% respectively. This implies that in the worst-case scenario the proposed enhance the accuracy by a factor of 12.62%, 15.92%, 27.67%, 16.68%, and 34.68% compared to the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS. While taking the best-case scenarios into account, the proposed ASM-RF improve the accuracy by 4.62%, 7.02%, 21.52%, 11.52%, and 29.18% when compared to the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS.

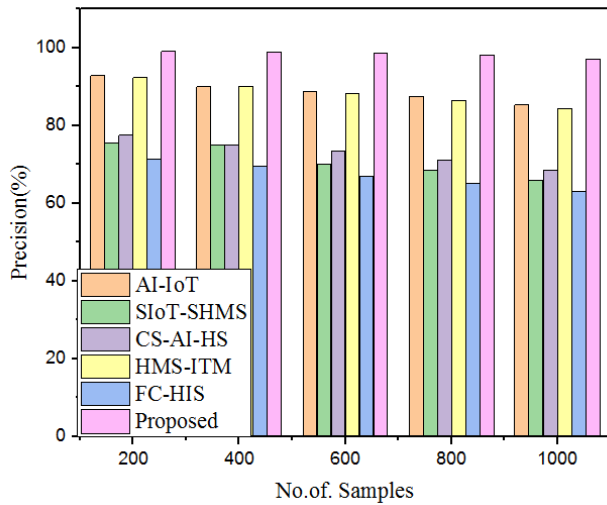


FIGURE 5. A comparison of precision (measured in percentage) of the proposed ASM-RF against AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS with varying numbers of sample data points.

2) PRECISION

The accuracy metric assesses a model’s performance by revealing how accurately it forecasts the actual labels for a given batch of data. It is determined by dividing the number of true positives (labels that were successfully predicted) by the total number of predictions tries and learning, and it aids in assessing how well a classification model performs. Consequently, the efficiency of the developed method has been determined by counting the amount of exact positive estimation methods that believe in true positive estimates as presented in Eq. (12).

$$pr = \frac{\vec{tp}}{\vec{tp} + \vec{fp}} \tag{12}$$

Considering the 200 data samples, the precision is 92.9%, 75.60%, 77.60%, 92.40%, 71.34%, and 99.12% with AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, FC-HIS, and proposed ASM-RF, respectively. This implies that the AI-IoT and HMS-ITM follow the proposed ASM-RF and perform well compared to the CS-AI-HS, SIoT-SHMS, and FC-HIS approaches. Following that, while considering the 1000 data samples the precisions with AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, FC-HIS, and proposed ASM-RF, are 85.30%, 66.00%, 68.45%, 84.50%, 63.03%, and 97.12%, respectively. A detailed comparison of the precision concerning the varying data samples is presented in Figure 5.

3) RECALL

Recall is a metric that measures how successfully a model can locate all relevant elements in a dataset. It is derived by dividing the number of true positives (correctly predicted labels) by the total number of actual positives (true labels in the dataset). The recall in our simulation is defined as the percentage of predictions that accurately identified the

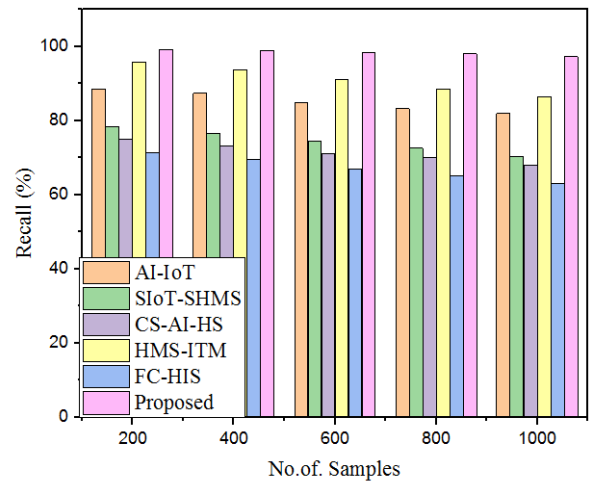


FIGURE 6. A comparison of recall (measured in percentage) of the proposed ASM-RF against AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS with a varying number of sample data points.

patient’s health state and is determined using Eq. (13).

$$re = \frac{\vec{re}}{\vec{tp} + \vec{fp}} \tag{13}$$

The recall performance assessment of the proposed ASM-RF against the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS methods is presented in Figure 6. It is evident from the figure that the proposed ASM-RF performs well compared to these competitive algorithms. In more detail for the 200 data samples, the recall is 88.60%, 78.40%, 75.00%, 95.90%, 71.34%, and 99.22% with the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, FC-HIS, and ASM-RF algorithms, respectively. While it is 82.00%, 70.24%, 68.00%, 86.50%, 63.03%, and 97.32%, with the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, FC-HIS, and ASM-RF algorithms, respectively. Consequently, the recall is significantly improved with the proposed ASM-RF algorithms when compared to the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS algorithms.

4) F-SCORE

The F-score is a measure of a model’s efficacy that is derived as the harmonic mean of precision and recall and offers an overall score of how well the model is doing as specified in Eq. (14). The F-score ranges from 0 to 1, with 1 being perfect (100%) and 0 being entirely incorrect (0%).

$$fm = 2 \left(\frac{pr \times re}{pr + re} \right) \tag{14}$$

where the pr represents the measured precision value and the re denotes the measured recall value. The F-score of the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, FC-HIS, and proposed ASM-RF is presented in Figure 7 for varying numbers of data samples. It can be observed that the F-score decreases with increasing the number of data samples, however, the proposed ASM-RF is still able to maintain the F-score by enhancing

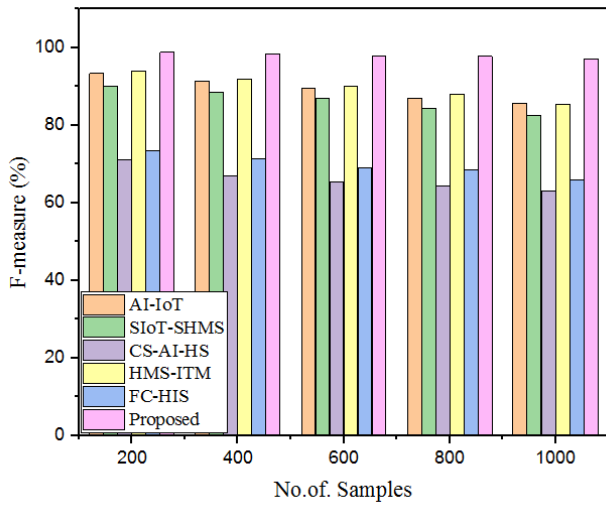


FIGURE 7. A comparison of F-score (measured in percentage) of the proposed ASM-RF against AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS with varying number of sample data points.

the performance compared to the other algorithms. Given the lowest number of data samples (200 data samples), the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, FC-HIS, and proposed ASM-RF has the F-score 93.40%, 90.10%, 71.10%, 94.00%, 73.53%, and 98.84%, respectively. While considering the maximum data samples (1000 data samples) then it is 85.78%, 82.50%, 62.98%, 85.50%, 66.00%, and 97.12% with the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, FC-HIS, and proposed ASM-RF.

5) EXECUTION TIME

The execution time is the amount of time it takes for an algorithm to run and can be estimated by counting the number of steps required to execute the algorithm and then multiplying the number of steps by the time each step takes, as shown in Eq. (15).

$$E(t) = I_c \times C_p \times C_t \tag{15}$$

where I_c defines the instruction count, C_p represents the cycles per instruction, and C_t denoted the clock cycle time. The execution time for the maximum number of data samples is considered and a comparison of the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, FC-HIS, and proposed ASM-RF is presented in Figure 8.

The proposed method exhibits an impressive execution time of 6 seconds, significantly outperforming several conventional AI-IoT methods such as HMS-ITM, CS-AI-HS, SIoT-SHMS, and FC-HIS, which require 25 seconds, 10 seconds, 12 seconds, 18 seconds, and 19 seconds respectively. This noteworthy outcome clearly highlights the time efficiency of the proposed method compared to traditional AI-IoT approaches, which not only consume more time for prediction and security tasks but also pose potential security risks. The extended execution time in conventional methods can create vulnerabilities in the healthcare system,

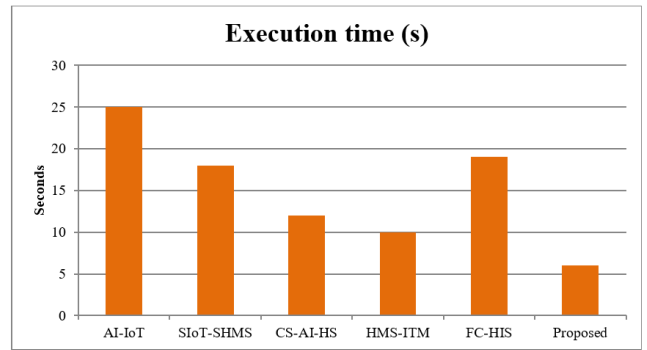


FIGURE 8. A comparison of execution time (measured in seconds) of the proposed ASM-RF against AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS with the maximum number of sample data points.

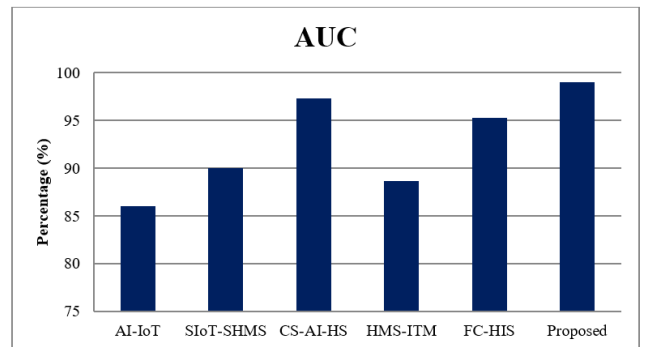


FIGURE 9. A comparison of area under the curve (measured in percentage) of the proposed ASM-RF against AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS.

enabling hackers to potentially access data before encryption at the user’s end. As, increasing the execution time not only compromises security but also adversely affects the overall effectiveness and cost of medical treatment. However, the comparative analysis reveals that the proposed ASM-RF secured disease prediction method achieves significantly lower execution times than its predecessors. This accomplishment underscores the superior efficiency and performance of the proposed approach.

6) AREA UNDER CURE

It characterizes the receiver operating characteristic (ROC) curve by representing the classifier’s ability to discriminate across classes. The effectiveness of the AUC is determined by the positive and negative classification of patient care and relates to the amount or degree of interpretability; a high AUC rate indicates a more accurate diagnosis of the condition. A comparison of the area under curve of the proposed ASM-RF against AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS is illustrated in Figure 9. It can be observed that the proposed ASM-RF acquire a higher AUC of 99.00%, followed by CS-AI-HS, FC-HIS, SIoT-SHMS, HMS-ITM, and AI-IoT with achieving AUCs 97.30%, 95.30%, 90.00%, 88.70%, and 86.00%, respectively.

TABLE 1. A comparative analysis of the various research studies in the healthcare system.

References	Accuracy (%)	Remote Access	Applications	Scalability Problem	Merits	Demerits
Amit Kishor and Chinmay [35]	97.62	Yes	Disease prediction	Yes	Different diseases are considered for validation	Accuracy is low
Hesham and Mustafa [36]	-	Yes	Hospital data security	No	Reliable, accurate and real-time monitoring	More error is possible for the execution
Taher M. Ghazal [37]	98.4	Yes	Healthcare security	Yes	Different diseases are considered for validation	Computational time is high
Reza et al. [39]	-	Yes	Secure healthcare	Yes	Human interaction can be reducing	Data privacy is the major issue
Naveed et al. [39]	-	Yes	Applications for health monitoring and other vital duties	No	Effective for important task applications, although the prediction module should be improved.	Cost reduction is not mentioned
A. Halder et al. [40]	95.75	No	COVID-19 detection	Yes	Simplisity	Lack of validation performance
H.P. Bui et al. [41]	-	No	Needle insertion simulation	Yes	Complexity is reduced	Lower accuracy
X. Zhang et al. [42]	94.46	No	COVID-19 detection	No	Less amount of parameter is used for the analysis	Misclassification rate is high
Ali et al. [43]	98.50	No	Heart disease prediction	No	Effective prediction is achieved	High complexity
Ali et al. [44]	90.00	No	Healthcare monitoring diabetics and BP patients	No	Enhanced the accuracy of the health condition	Lack of efficient performance metrics
Shaker El-Sappagh [45]	86.17	No	Alzheimer's disease detection	No	Inexpensive and extensively used for different application	Complex to decision making process
ASM-FR Hybrid Framework	99.52	Yes	Cancer, cardiovascular detection	Yes	High predictive and security performance	Need to varification large datasets

B. DISCUSSIONS

The proposed approach's performance is compared with state-of-the-art methods, showcasing its superiority in Table 1. This comprehensive table highlights the advancements and excellence of the proposed approach across various performance metrics. It is evident that the majority of previous studies lack the inclusion of the created model as an integral part of a comprehensive framework, limiting its application and accessibility. In contrast, the effectiveness of the proposed work demonstrates its potential for utilization in remote areas, enabling healthcare providers to serve patients located far away through IoT-based mobile applications. This innovative approach not only strengthens the connection between patients and healthcare providers but also extends healthcare services to underserved regions, thereby enhancing healthcare access and delivery.

In comparison to the AI-IoT, SIoT-SHMS, CS-AI-HS, HMS-ITM, and FC-HIS algorithms, the proposed ASM-RF technique demonstrates significant advancements in multiple evaluation metrics, including accuracy, precision, AUC,

F-score, and recall, while effectively addressing challenges encountered during the training process. By employing a strategic approach, early training issues were successfully resolved, leading to substantial improvements in data categorization across diverse data sample points and consequently enhancing the prediction accuracy of the model. The notable performance enhancement of the ASM-RF stems from its ability to augment the model with a larger number of decision trees, facilitating more comprehensive search capabilities while optimizing computational resources and reducing runtime. This feature empowers the ASM-RF to provide more reliable and accurate forecasts, thereby elevating the overall performance of the Random Forest model. Specifically, the results obtained from the proposed ASM-RF are highly impressive. The technique consistently achieved exceptional performance metrics, including a remarkable AUC of 99.00%, recall of 99.22%, precision of 99.12%, and accuracy of 99.52%. Furthermore, these outstanding outcomes were accomplished within a remarkably short execution time of only 6 seconds, highlighting the efficiency

and speed of the proposed approach. The ASM-RF not only outperforms other algorithms in terms of accuracy and precision but also showcases its robustness and resilience in accurately diagnosing the health status of patients. The proposed ASM-RF serves as a highly reliable and effective tool for supporting precise and accurate health status assessments, making it a valuable contribution to the field of disease detection and diagnosis.

The efficacy of the proposed work demonstrates its potential for facilitating remote patient monitoring through the utilization of IoT-based mobile applications. This innovative approach empowers clinicians and healthcare professionals to deliver treatment and provide invaluable recommendations to patients in a manner that is not only reliable and secure but also remarkably user-friendly. By leveraging the capabilities of IoT technology, this framework facilitates seamless communication and data exchange, streamlining the healthcare process and enhancing patient care. Its deployment holds immense promise for transforming the healthcare landscape, enabling practitioners to offer personalized and efficient medical interventions while ensuring patient well-being.

C. LIMITATION OF THE STUDY

Despite implementing the proposed framework on a considerable dataset, it is crucial to acknowledge a significant limitation in this study: the utilization of a relatively small-scale dataset. This constraint hampers the generalizability and application of the findings to more extensive and heterogeneous healthcare scenarios. Furthermore, conducting thorough security assessments is imperative to validate the efficacy of the employed encryption scheme and fortify data protection, especially when dealing with larger-scale datasets. Additionally, to ascertain the true potential and comparative prowess of the ASM-RF framework, it is imperative to conduct additional benchmarking against a diverse range of well-established algorithms and expansive datasets. By addressing these limitations, we can augment the credibility and broad applicability of the proposed framework within real-world healthcare settings.

VI. CONCLUSION

In this paper, we developed a novel ASM-RF hybrid framework that leverages the ASM algorithm to construct the trees and pairs the predictive analytics of a Random Forest algorithm to evaluate the pattern of health data to spot out and identify modest indications, enhancing the computational efficiency and accuracy. The ASM-RF developed utilizes the fitness function to evaluate the spider monkey's performance at the classification layer, influencing accuracy and recall, resulting in more accurate patient disease diagnoses, and automating timely treatment decisions to enhance the overall healthcare system. Besides taking into consideration the relevance of the sensitivity of the healthcare data, identity-based encryption is used, which enables the encryption of data using private and public keys coupled with the user's identities as encryption keys, thereby contributing to increasing the

security of the healthcare system. The proposed ASM-RF hybrid framework is simulated using a dataset gathered from Kaggle, comprising data from three distinct IoT sensor devices, ten participants, and twelve activities. This framework is subsequently compared to state-of-the-art predictive diagnostic algorithms, evaluating its performance based on accuracy, precision, Area under the Curve, execution time, recall, and F-measure. When compared to cutting-edge diagnosis algorithms, the findings showed high accuracy (99.52%) and minimal execution time (6 seconds).

In light of the limitations discussed earlier, our future work will primarily focus on conducting a comprehensive evaluation of the proposed framework. Our main objective is to extensively explore and analyze its performance by subjecting it to rigorous benchmarking against a diverse range of well-established algorithms and utilizing expansive datasets. By undertaking a thorough comparison and evaluation against existing methods, we aim to gain a comprehensive understanding of the framework's strengths and weaknesses, identifying areas for further improvement and optimization. Expanding the scope of evaluation datasets will contribute to the robustness and generalizability of the framework, providing a more realistic representation of real-world scenarios and enhancing the reliability of results.

ACKNOWLEDGMENT

The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Research Priorities and Najran Research funding program. Grant Code (NU/NRP/SERC/12/11).

REFERENCES

- [1] S. Chatterjee, A. Pradhan, A. Shinde, and P. Verma, "Early diagnosis and prevention of diseases: A review," *Int. J. Preventive Med.*, vol. 6, no. 11, p. e100, 2015.
- [2] D. Kulkarni, A. Sagare, R. Patil, and P. Salunkhe, "Challenges of early disease detection and prevention," *Indian J. Public Health Res. Develop.*, vol. 8, no. 1, pp. 83–92, 2017.
- [3] L. O. Gostin, E. A. Friedman, and S. A. Wetter, "Responding to COVID-19: How to navigate a public health emergency legally and ethically," *Hastings Center Rep.*, vol. 50, no. 2, pp. 8–12, 2020.
- [4] M. K. Al-Hanawi, S. A. Khan, and H. M. Al-Borie, "Healthcare human resource development in Saudi Arabia: Emerging challenges and opportunities—A critical review," *Public Health Rev.*, vol. 40, no. 1, pp. 1–16, Dec. 2019.
- [5] S. Hermes, T. Riasanow, E. K. Clemons, M. Böhm, and H. Krmar, "The digital transformation of the healthcare industry: Exploring the rise of emerging platform ecosystems and their influence on the role of patients," *Bus. Res.*, vol. 13, no. 3, pp. 1033–1069, Nov. 2020.
- [6] H. Li and R. Cai, "Design and implementation of environmental monitoring system using photoelectric wireless sensor network," *J. Nanoelectron. Optoelectron.*, vol. 16, no. 10, pp. 1657–1666, Oct. 2021.
- [7] J. Hunter, "Patient centered strategy: A learning system for better care," in *Proc. Catalysis, Summit*, Chicago, IL, USA, 2018, p. 14-1.
- [8] R. Dias and A. Torkamani, "Artificial intelligence in clinical and genomic diagnostics," *Genome Med.*, vol. 11, no. 1, 2019, Art. no. 70.
- [9] A. Sharifi, M. Ahmadi, and A. Ala, "The impact of artificial intelligence and digital style on industry and energy post-COVID-19 pandemic," *Environ. Sci. Pollut. Res.*, vol. 28, no. 34, pp. 46964–46984, Sep. 2021.
- [10] K. Takleef, K. Ali, M. A. Salim, and M. Wadi, "An overview of patient's health status monitoring system based on Internet of Things (IoT)," *Wireless Pers. Commun.*, vol. 114, no. 3, pp. 2235–2262, 2020.

- [11] S. Wan, Z. Gu, and Q. Ni, "Cognitive computing and wireless communications on the edge for healthcare service robots," *Comput. Commun.*, vol. 149, pp. 99–106, Jan. 2020.
- [12] T. M. Ghazal, M. K. Hasan, M. T. Alshurideh, H. M. Alzoubi, M. Ahmad, S. S. Akbar, B. Al Kurdi, and I. A. Akour, "IoT for smart cities: Machine learning approaches in smart healthcare—A review," *Future Internet*, vol. 13, no. 8, p. 218, Aug. 2021.
- [13] M. Cubric, "Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study," *Technol. Soc.*, vol. 62, Aug. 2020, Art. no. 101257.
- [14] H. Bhardwaj, P. Tomar, A. Sakalle, and U. Sharma, "Principles and foundations of artificial intelligence and Internet of Things technology," in *Artificial Intelligence to Solve Pervasive Internet of Things Issues*. Amsterdam, The Netherlands: Elsevier, 2021, pp. 377–392.
- [15] M. D. Gurven and D. E. Lieberman, "WEIRD bodies: Mismatch, medicine and missing diversity," *Evol. Human Behav.*, vol. 41, no. 5, pp. 330–340, Sep. 2020.
- [16] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *J. Big Data*, vol. 6, no. 1, pp. 1–18, Dec. 2019.
- [17] B. Ihnaini, M. A. Khan, T. A. Khan, S. Abbas, M. S. Daoud, M. Ahmad, and M. A. Khan, "A smart healthcare recommendation system for multidisciplinary diabetes patients with data fusion based on deep ensemble learning," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–11, Sep. 2021.
- [18] A. Stanley and J. Kucera, "Smart healthcare devices and applications, machine learning-based automated diagnostic systems, and real-time medical data analytics in COVID-19 screening, testing, and treatment," *Amer. J. Med. Res.*, vol. 8, no. 2, pp. 105–117, 2021.
- [19] J. G. Richens, C. M. Lee, and S. Johri, "Improving the accuracy of medical diagnosis with causal machine learning," *Nature Commun.*, vol. 11, no. 1, p. 3923, Aug. 2020.
- [20] X. Sui, S. He, S. B. Vilsen, J. Meng, R. Teodorescu, and D.-I. Stroe, "A review of non-probabilistic machine learning-based state of health estimation techniques for lithium-ion battery," *Appl. Energy*, vol. 300, Oct. 2021, Art. no. 117346.
- [21] L. Lu, L. Xu, B. Xu, G. Li, and H. Cai, "Fog computing approach for music cognition system based on machine learning algorithm," *IEEE Trans. Computat. Social Syst.*, vol. 5, no. 4, pp. 1142–1151, Dec. 2018.
- [22] R. Vashistha, A. K. Dang, A. Kumar, D. Chhabra, and P. Shukla, "Futuristic biosensors for cardiac health care: An artificial intelligence approach," *3 Biotech*, vol. 8, no. 8, pp. 1–11, Aug. 2018.
- [23] M. A. Khan, "An IoT framework for heart disease prediction based on MDCNN classifier," *IEEE Access*, vol. 8, pp. 34717–34727, 2020.
- [24] G. S. Karthick and P. B. Pankajavalli, "Chronic obstructive pulmonary disease prediction using Internet of Things-spiro system and fuzzy-based quantum neural network classifier," *Theor. Comput. Sci.*, vol. 941, pp. 55–76, Jan. 2023.
- [25] N. V. L. M. K. Munagala, L. R. R. Langoju, A. D. Rani, and D. V. R. K. Reddy, "Optimised ensemble learning-based IoT-enabled heart disease monitoring system: An optimal fuzzy ranking concept," *Comput. Methods Biomech. Biomed. Eng. Imag. Vis.*, pp. 1–17, Jan. 2023.
- [26] R. Krishnamoorthi, S. Joshi, H. Z. Almarzouki, P. K. Shukla, A. Rizwan, C. Kalpana, and B. Tiwari, "A novel diabetes healthcare disease prediction framework using machine learning techniques," *J. Healthcare Eng.*, vol. 2022, pp. 1–10, Jan. 2022.
- [27] 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, pp. 122259–122269, 2020.
- [28] M. A. Khan, M. T. Quasim, N. S. Alghamdi, and M. Y. Khan, "A secure framework for authentication and encryption using improved ECC for IoT-based medical sensor data," *IEEE Access*, vol. 8, pp. 52018–52027, 2020.
- [29] S. Nandy, M. Adhikari, M. A. Khan, V. G. Menon, and S. Verma, "An intrusion detection mechanism for secured IoMT framework based on swarm-neural network," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 5, pp. 1969–1976, May 2022.
- [30] A. Verma, G. Agarwal, and A. K. Gupta, "A novel generalized fuzzy intelligence-based ant lion optimization for Internet of Things based disease prediction and diagnosis," *Cluster Comput.*, vol. 25, no. 5, pp. 3283–3298, Oct. 2022.
- [31] M. K. Ahirwar, P. K. Shukla, and R. Singhai, "CBO-IE: A data mining approach for healthcare IoT dataset using chaotic biogeography-based optimization and information entropy," *Sci. Program.*, vol. 2021, pp. 1–14, Oct. 2021.
- [32] N. Khare, P. Devan, C. Chowdhary, S. Bhattacharya, G. Singh, S. Singh, and B. Yoon, "SMO-DNN: Spider monkey optimization and deep neural network hybrid classifier model for intrusion detection," *Electronics*, vol. 9, no. 4, p. 692, Apr. 2020.
- [33] A. L. H. P. Shaik, M. K. Manoharan, A. K. Pani, R. R. Avala, and C.-M. Chen, "Gaussian mutation-spider monkey optimization (GM-SMO) model for remote sensing scene classification," *Remote Sens.*, vol. 14, no. 24, p. 6279, Dec. 2022.
- [34] M. K. Choudhary and S. Hiranwal, "Feature selection algorithms for plant leaf classification: A survey," in *Proc. Int. Conf. Commun. Comput. Technol. (ICCCCT)*. Singapore: Springer, 2021, pp. 657–669.
- [35] A. Kishor and C. Chakraborty, "Artificial intelligence and Internet of Things based healthcare 4.0 monitoring system," *Wireless Pers. Commun.*, vol. 127, no. 2, pp. 1615–1631, Nov. 2022.
- [36] H. A. El Zouka and M. M. Hosni, "Secure IoT communications for smart healthcare monitoring system," *Internet Things*, vol. 13, Mar. 2021, Art. no. 100036.
- [37] T. M. Ghazal, "Internet of Things with artificial intelligence for health care security," *Arabian J. Sci. Eng.*, vol. 48, no. 4, p. 5689, Apr. 2023.
- [38] M. R. Davahli, W. Karwowski, K. Fiok, T. Wan, and H. R. Parsaei, "Controlling safety of artificial intelligence-based systems in healthcare," *Symmetry*, vol. 13, no. 1, p. 102, Jan. 2021.
- [39] Imran, N. Iqbal, S. Ahmad, and D. H. Kim, "Health monitoring system for elderly patients using intelligent task mapping mechanism in closed loop healthcare environment," *Symmetry*, vol. 13, no. 2, p. 357, Feb. 2021.
- [40] A. Halder and B. Datta, "COVID-19 detection from lung CT-scan images using transfer learning approach," *Mach. Learn., Sci. Technol.*, vol. 2, no. 4, Dec. 2021, Art. no. 045013.
- [41] H. P. Bui, S. Tomar, and S. P. A. Bordas, "Corotational cut finite element method for real-time surgical simulation: Application to needle insertion simulation," *Comput. Methods Appl. Mech. Eng.*, vol. 345, pp. 183–211, Mar. 2019.
- [42] X. Zhang, J. Zhou, W. Sun, and S. Kumar Jha, "A lightweight CNN based on transfer learning for COVID-19 diagnosis," *Comput., Mater. Continua*, vol. 72, no. 1, pp. 1123–1137, 2022.
- [43] F. Ali, S. El-Sappagh, S. M. R. Islam, D. Kwak, A. Ali, M. Imran, and K.-S. Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Inf. Fusion*, vol. 63, pp. 208–222, Nov. 2020.
- [44] F. Ali, S. El-Sappagh, S. M. R. Islam, A. Ali, M. Attique, M. Imran, and K.-S. Kwak, "An intelligent healthcare monitoring framework using wearable sensors and social networking data," *Future Gener. Comput. Syst.*, vol. 114, pp. 23–43, Jan. 2021.
- [45] S. El-Sappagh, F. Ali, T. Abuhmed, J. Singh, and J. M. Alonso, "Automatic detection of Alzheimer's disease progression: An efficient information fusion approach with heterogeneous ensemble classifiers," *Neurocomputing*, vol. 512, pp. 203–224, Nov. 2022.
- [46] S. Bouktif, A. Fiaz, A. Ouni, and M. Serhani, "Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches," *Energies*, vol. 11, no. 7, p. 1636, Jun. 2018.
- [47] V. Vijayakumar, D. Malathi, V. Subramaniyaswamy, P. Saravanan, and R. Logesh, "Fog computing-based intelligent healthcare system for the detection and prevention of mosquito-borne diseases," *Comput. Hum. Behav.*, vol. 100, pp. 275–285, Nov. 2019.



REYAZUR RASHID IRSHAD received the B.Sc. degree from Aligarh Muslim University, Aligarh, India, in 2000, and the master's degree in computer application from Indira Gandhi University, New Delhi, India, in 2010. He is currently pursuing the Ph.D. degree with JJT University, Rajasthan. He is a Lecturer with the Department of Computer Science, Najran University, Saudi Arabia. He has published many articles in reputed journals and has attended some conferences. His research interest includes web-based applications.



SHAHID HUSSAIN received the B.S. degree in mathematics and the M.Sc. degree in computer science from the University of Peshawar, in 2002 and 2005, respectively, and the M.S. and Ph.D. degrees in computer engineering from Jeonbuk National University, South Korea, in 2016 and 2020, respectively. He was a Postdoctoral Researcher with the Gwangju Institute of Science and Technology (GIST), South Korea, in 2020, and the University of Galway (UoG), Ireland, from 2020 to 2022.

He is currently a Senior Postdoctoral Researcher with the Innovation Value Institute (IVI), School of Business, National University of Ireland Maynooth (NUIM), Ireland. His research interests include smart grids, energy management, electric vehicles, smart grid infrastructure, optimization algorithms, micro-grid operations, distributed energy resources, peer-to-peer energy trading, and machine learning in medical applications (e.g., prediction and risk analysis of osteoporosis) using fuzzy logic, game theory, ontology, AI, and blockchain approaches and technologies. He received the Jeonbuk National University Presidential Award for academic excellence during the Ph.D. studies.



IHTISHAM HUSSAIN is currently pursuing the B.S. degree in computer science with the Abdul Wali Khan University of Mardan, Khyber Pakhtunkhwa, Pakistan. He has participated in several coding competitions, hackathons, internships, and projects. Besides his academic interests, he is a devoted programmer and a technology enthusiast, continuously investigating new tools and approaches to enhance his coding talents. His research interests include artificial intelligence

and machine learning applications in a variety of fields, such as bioinformatics, energy, and business.



AHMED ABDU ALATTAB received the B.Sc. degree in computer science from the University of Baghdad, Baghdad, Iraq, in 1997, the M.Sc. degree in computer science from the University of Technology, Baghdad, in 2002, and the Ph.D. degree in computer science-artificial intelligence from the University of Malaya, Kuala Lumpur, Malaysia, in 2013. He is currently an Assistant Professor with the Department of Computer Science, Najran University, Saudi Arabia. He is also an Assistant

Professor with Tamar University, Yemen. He has published many papers in reputed journals and conferences. His research interests include artificial intelligence, image processing, image retrieval, image classification, object recognition, deep learning, natural language processing, computer vision, and neural networks.



ADIL YOUSIF received the B.Sc. and M.Sc. degrees from the University of Khartoum, Sudan, and the Ph.D. degree from the University of Technology in Malaysia (UTM). He is currently an Associate Professor with the College of Arts and Sciences Sharourah, Najran University, Saudi Arabia. He is also the principal investigator of several research projects in artificial intelligence and emerging technologies. His research interests include computer networks, cloud computing, artificial intelligence, and optimization techniques.



OMAR ALI SALEH ALSAIARI received the B.Sc. degree in computer science from Tabuk University, Tabuk, Saudi Arabia, in 2008, and the M.Sc. degree in computer science from the University of New Brunswick, Fredericton, U.K., in 2013. He is currently a Lecturer with the Department of Computer Science, Najran University, Saudi Arabia. His research interests include artificial intelligence, software engineering, and data mining.



ELSHAREEF IBRAHIM IDREES IBRAHIM received the B.Sc. degree from the Department of Mathematics, Faculty of Science, Nilein University, in 2003, the M.Sc. degree from the Department of Mathematics, Faculty of Science, Khartoum University, Sudan, in 2009, and the Ph.D. degree in pure mathematics-analysis (pure and applied approximation theory) from the Department of Mathematics, Faculty of Science, University of Sudan, in 2019. He joined Najran

University, as an Academic Staff, in 2011, after completing the M.Sc. degree in pure mathematics. His research interests include foundations of mathematics analysis, skills and expertise, pure mathematics, mathematical analysis, set theory, functional analysis, real analysis, group theory, Fourier analysis, matrix theory, differentiation numbers, integration, and complex analysis.

...