International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No.1 (Sep-24)

http://dx.doi.org/10.12785/ijcds/160182

Diabetic Patient Real-Time Monitoring System Using Machine Learning

Tariq Emad Ali[1], Faten Imad Ali[2], Ameer Hussein Morad[3], Mohammed A. Abdala[4] and Alwahab Dhulfiqar Zoltan[5]

[1]*Department of Information and Communication Engineering, Al-Khwarizmi College of Engineering, University of Baghdad, Baghdad, Iraq*
[2]*Department of Biomedical Engineering, College of Engineering, AL-Nahrain University, Baghdad, Iraq*
[3]*College of Engineering Technology, Gilgamesh University, Baghdad, Iraq*
[4]*Head of Medical Instruments Techniques Engineering Department, Al-Hussain University College, Kerbala, Iraq*
[5]*Faculty of Informatics, Eotvos Lorand University, Budapest, Hungary*
[1]*Received 19 Feb. 2024, Revised 20 May 2024, Accepted 21 May 2024, Published 1 Sep. 2024*

Abstract: Continuous monitoring is critical to improving the quality of life of people with diabetes. Leveraging technologies such as the Internet of Things (IoT), modern communication tools, and artificial intelligence (AI) can contribute to reducing healthcare costs. The integration of various communication systems allows the provision of personalized and remote healthcare services. The increasing volume of healthcare data poses challenges in storage and processing. To overcome this challenge, this paper suggests intelligent medical architectures for intelligent e-health applications. To provide cutting-edge medical services, 5G and 6G technologies are necessary, since they can satisfy critical needs, including high bandwidth and energy efficiency. This work presents an intelligent machine learning (ML) using an ensemble learning-based real-time monitoring system for diabetes patients. Mobiles, detectors, and other intelligent gadgets are used as buildings to gather measurements of the body. Subsequently, the collected data undergoes a normalization procedure for preprocessing. Principal Component Analysis (PCA) is employed to extract features. The ranking of every feature in the dataset is then assessed using two feature selection (FS) techniques, namely information gain (IG) and chi-square (chi2), and the association between the features chosen by the FS methods is then found using Pearson's correlation method, which is one of the correlation methods that can be used to find the correlated between the selected features. For diagnostic purposes, the intelligent system employs data classification through an ensemble learning approach using XGBoost and Random Forest (RF) as base models, which is named (ENS_XGRF). The final classification is determined by a hard voting mechanism in conjunction with particle swarm optimization (PASWOP). The simulation results underscore the superiority of the suggested approach in terms of accuracy when compared to alternative techniques.

Keywords: Internet of Things, Machine Learning, Principal Component Analysis, Particle Swarm Optimization.

1. INTRODUCTION

The healthcare sector is always growing, which creates a lot of study problems in the field of information technology. These problems can be successfully addressed by combining developments in detectors, ML, AI, big data analysis, and information and communication technology (ICT). Especially, IoT-enabled message monitoring equipment allows clients to predict medical issues like high temperatures, diabetes issues, heart attacks, etc., which helps with healthy living, assisted care for the elderly, and preventative therapy [1]. This approach not only provides reliable assistance but also mitigates travel issues of patients, thus improving the overall quality of care. The advent of new technologies is primarily geared towards the continuous monitoring of patients with prolonged diseases, a prevalent trend in current times [2]. Chronic diseases, which are defined by their long course and continuous care requirements, frequently require prolonged hospital admissions to provide periodic assessments. Chronic diseases include diseases such as diabetes, cancer, and heart disease; diabetics have become a prevalent disease, as well as an extremely severe disease that takes many lives annually [2]. Since it is a chronic disease caused by pancreatic failure, frequent and rigorous treatment is essential to avoid damage to different parts such as the veins, nerve cells, and eyelids [3]. The surge in diabetic patients underscores the necessity for advanced technologies to monitor and manage their health effectively [4]. Routine monitoring of blood glucose levels is standard practice for individuals with diabetes, allowing for continuous tracking by diabetics, their families, and medical professionals [5]. Portable surveillance systems offer the advantage of reducing hospital stays, thereby improving the quality of life for

E-mail address:tariqemad@gmail.com,faten.imad@nahrainuniv.edu.iq,ameer$_m$orad@yahoo.com

diabetic patients [6]. To facilitate seamless communication of patient information to healthcare providers, the adoption of 5G and 6G technology is proposed, offering high-speed communication and improved network adaptability, such as bandwidth and capacity [7]. Despite the abundance of recent research proposing intelligent patient tracking systems, many do not address the critical aspects of accurate, timely assistance and the delay-associated parameters of measured parameters during transmission [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], and [19]. The problem statement revolves around enhancing the quality of life for individuals with diabetes through continuous monitoring, and utilizing modern technologies such as the IoT, communication tools, and AI. Key points include:

- Managing Healthcare Data Challenges: The increasing volume of healthcare data presents storage and processing challenges that require intelligent architectures for effective management.
- Utilizing 5G and 6G Technologies: The integration of 5G and 6G technologies is proposed to meet critical requirements such as high bandwidth and energy efficiency, supporting advanced medical services.
- Developing Intelligent Monitoring Systems: The paper proposes an intelligent machine learning-based real-time monitoring system for diabetic patients that uses mobile devices, detectors, and other gadgets for data collection. This data is subjected to preprocessing and feature extraction using techniques such as normalization and PCA.
- Employing Feature Selection and Correlation Analysis: The extracted features are selected using information gain (IG) and chi-square (chi2) techniques, with the association between the features evaluated using Pearson's correlation method.
- Conducting Diagnostic Classification with Ensemble Learning: Data classification for diagnostic purposes employs an ensemble learning approach, specifically XGBoost and Random Forest (RF) as base models. The final classification utilizes a hard voting mechanism combined with particle swarm optimization (PASWOP).
- Demonstrating the Superiority of Approach: The simulation results demonstrate the superiority of the proposed approach in terms of accuracy compared to alternative techniques.

In summary, the problem statement addresses the necessity for advanced technologies and intelligent systems to enhance diabetes management through continuous monitoring and diagnostic classification, aiming for improved healthcare outcomes. In response, this study introduces an IoT-based health service that integrates ML algorithms to provide proactive real-time assistance during medical emergencies. The main goal is to use ML to predict and connect gadgets over the Internet IoT, to gather ongoing historical information, and interpret it in the cloud.

This study contributes primarily to the field by utilizing ML to build an intelligent technique for real-time tracking of people with diabetes. The structural elements of the suggested framework consist of smartphones, electronic gadgets, and detectors that are all used to get important human measures. The information collected is then preprocessed using a standardization technique. To extract features from the preprocessed data, Principal Component Analysis (PCA) is employed. Finally, the core of the functionality of the intelligent system lies in the data categorization process. For this purpose, an innovative approach is introduced that combines XGBoost and Random Forest (RF) as base models, and the final classification is determined by a hard voting mechanism in conjunction with PASWOP and the proposed model named ensemble methods with XGBoost and Random Forest (ENS_XGRF). This integrated methodology is used to generate a diagnosis, providing valuable information on the health status of patients with diabetes. The paper is structured into four main parts: Part (2) focuses on the related work; Part (3) outlines the methodology of the proposed framework; Part4 presents the study results; and Part 5 provides the conclusion.

2. Related Work

This part presents a wide range of research investigations on diabetes sufferers' wellness tracking methods.

The authors of [8] addressed issues with ongoing surveillance, the lack of anomaly identification techniques, and the requirement of long training times for forecasting strategies by introducing a dynamic and context-aware tracking platform. prediction. In [9], technological developments are described for creating an architecture for safeguarding information from the health Internet of Things (HIoT) together with a safety analysis. Through a variety of use-case situations, this study provides an organizational structure for monitoring wellness indicators in individuals with impairments or chronic progressive illnesses. Several algorithms were used in a methodical investigation in [10] to comprehend classifiers for determining the prevalence of type-A diabetes in humans. A methodology to make fast and reliable disease predictions is called an ”intelligent medical referral framework for patients with multifunctional diabetes” and is provided in [11]. However, the need for a more comprehensive, effective, diagnostic and recommendation method is placed on a wide range of human diseases. [12] provides a detailed analysis of ubiquitous, intelligent, and connected medical facilities to monitor people with chronic and lifestyle conditions. Deep learning (DL) and cloud-based analytics are used in the design to provide intelligent patient surveillance and control. The study described in [13] utilizes ML-SVM to forecast the probability of diabetes. The approach focuses specifically on women within the dataset who share a Pima Indian heritage. The [14] is focused on immediate information to improve forecasting and accuracy through the use of ML and IoT, together with a recommended software and hardware solution to support the early detection of heart disease. [15] discusses opposing cutting-edge health care for elderly patients and their caregivers. Although acknowledging many overlooked achievements in the area, [16] performs a thorough evaluation of approaches for the diagnosis, recognition, and management of diabetes mellitus. [17] introduces a novel approach to tracking one's health that utilizes a safe information storage structure for patient information in cloud-based platforms in conjunction with main information collected from folks in remote areas to anticipate diseases. [18] describes the creation and creation of a software platform that uses ML to improve adherence to therapy. The system of monitoring suggested in [20] addresses the effects of factors on the health of diabetic patients. The paper in [21] examines prospects for expansion in the sector and discusses the advantages of combining AI with telehealth. [22] uses supervised ML classification techniques to predict hypertension and diabetes conditions based on patient sugar and arterial pressure information. [23] proposes a light-transmission model that uses Li-Fi technology to determine the body's glucose levels. [24] uses AI and IoT to research the medical industry to improve patient care and assistance. The study by [25] presents a shallow neural network with immediate data in IoT within the intelligent medical strategy. The technology that goes into creating 5G e-health services is covered within [26] through a variety of angles. Table I shows the literature survey.

3. Proposed Framework

In the construction of a classification using Machine Learning (ML), the fundamental processes depicted in Figure 1 include preprocessing, feature extraction, and classification.

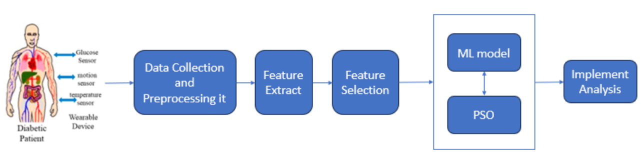

Figure 1. Proposed framework.

In the initial stages of building the framework, Python serves as the primary language for both data preprocessing and exploration. Libraries such as Pandas, NumPy, and scikit-learn are utilized for tasks like data cleaning, feature engineering, and statistical analysis. Following this, the machine learning model development phase leverages TensorFlow, Keras, and scikit-learn to build and train predictive models using the preprocessed data. Real-time data collection and processing are facilitated through wearable sensors and glucose monitors, enabling continuous monitoring of patient data. Visualization of insights is accomplished using Plotly, Matplotlib, and Seaborn, allowing for the creation of informative and interactive visualizations to aid in data exploration and model interpretation. Overall, this framework integrates various tools to streamline the processes of data preprocessing, model development, real-time data collection, and visualization for effective diabetes monitoring. The study utilizes ensemble learning for intelligent patient monitoring, combining XGBoost and Random Forest models. Preprocessing involves normalization and feature extraction using PCA and chi-square data with information gain. Pearson's correlation assesses feature correlation. The final classification uses hard voting and particle swarm optimization. The method outperforms current techniques across various metrics, emphasizing scalability and user privacy.

A. Data Collection and Preprocessing

In this study, the dataset includes the records of 62 individuals with diabetes, consisting of 44 males and 18 females, who underwent an average of 67 days of examinations. The diabetes dataset [27] was constructed using data collected from the Iraqi society, sourced from the laboratories of Medical City Hospital and the Specialized Center for Endocrinology and Diabetes at Al-Kindy Teaching Hospital. Patient files were obtained, and relevant information was extracted and entered into the database. The dataset comprises medical information and laboratory analyses, including blood sugar levels, age, gender, creatinine ratio (Cr), body mass index (BMI), urea, cholesterol (Chol), fasting lipid profile (total, LDL, VLDL, triglycerides, and HDL cholesterol), HbA1C, and diabetes disease class (diabetic, non-diabetic, or predict-diabetic). The dataset related to glucose concentration comprises a total of 12,612 data points, where each data point is characterized by 5 features. Before applying ML algorithms, a crucial preliminary step is data preprocessing (i.e., data preparation). This is necessary because real-world data often exhibit noise, insufficiency, and unreliability, making them unsuitable for immediate use in the prediction process. Preparation is considered necessary to use data for the diagnosis of diabetic disease in an efficient manner. Data preparation is difficult because every diabetic illness record denoted as Di^{il}, has a range of features, every single one of which is characterized by a distinct set of numerical values. To address this difficulty, a normalization mechanism has been placed in effect to scale the value Di^{il} between 0 and 1, which lowers the computational complexity involved in determining the prediction of diabetes. Among the various techniques available for data normalization, the suggested system employs a min–max normalization method.

B. Feature Extraction

Feature extraction poses a significant challenge in machine learning, underscoring the importance of generating new dimensions by combining existing ones. Principal Component Analysis (PCA) emerges as a pivotal technique in addressing this challenge. PCA works by transforming dimensions into a new set of dimensions termed principal components. These components are eigenvectors linked to the highest eigenvalues of the covariance matrix, effectively

TABLE I. Literature survey

Ref. No.	Title	Author	Advantages	Disadvantages
[8]	A Storage Optimization and Energy Efficiency-Based Edge-Enabled Companion-Side eHealth Monitoring System for IoT-Based Smart Hospitals	Gharaei et.al., 2023	Optimizes storage, energy efficiency, cost savings; reduces data latency, consumption. Enables real-time companion-side monitoring, and personalized interventions in smart hospitals.	Data security, privacy risks. Edge device limitations hinder analytics. Maintenance and interoperability issues complicate IoT deployment.
[9]	IoT in E-Health, Assisted Living, and E-Wellness	Azim et.al., 2023	Remote glucose monitoring, self-management via smart devices, improved glycemic control with IoT insulin delivery systems	Privacy concerns, reliability issues with IoT devices, and potential inaccuracies in data interpretation.
[10]	Comparative Approach for Early Diabetes Detection with Machine Learning.	Harnal et al. 2023	SVMs are good at handling high-dimensional data and small datasets.	Unsuitable for large data sets and long training time.
[12]	Novel framework based on deep learning and cloud analytics for smart patient monitoring and recommendation	Motwani et.al.,2023	using DL make it is robust to outliers	A lot of training data is needed
[13]	A Diabetes Monitoring System and Health Medical Service Composition Model in Cloud Environment	Sharma et.al. 2023	cost savings, reduced data transfer latency, lower energy consumption and enabling real-time access to patient health data	Risks associated with data security and privacy.
[14]	Smart Healthcare Monitoring System Using IoT	Mohammed et.al.,2023	Support Vector Machine it is robust to outliers	Unsuitable to Large Datasets
[?]	Hybrid Adaptive Machine Learning Approach for Intelligent Patient Monitoring in e-Health Systems	Dohare et.al.,2024	Innovative, high accuracy, integrates cutting-edge tech, hybrid architecture	Limited sample, complex implementation, potential overfitting
[21]	Clinical Prediction on ML-based Internet of Things for E-Health Care System	Chaturvedi et al.and, 2023	ANFIS handles complex, nonlinear relationships in diabetes monitoring, providing accurate predictions, personalized recommendations. Adapts improve with new data.	Expertise needed for ANFIS modeling, the potential for overfitting, risk of inaccurate predictions, and patient harm.
[23]	Development of A Smart Non-Invasive Glucose Monitoring System With SpO2 and BPM for Diabetic Patients	Nabil et al., 2023	Non-invasive, quick glucose, SpO2, heartbeat measurements. Convenience, real-time monitoring, LCD, reduced discomfort, no blood samples.	Less scalable and adaptable compared to ML. Relies on predetermined algorithms and lacks flexibility for individual variations
[24]	Artificial intelligence in healthcare delivery: Prospects and pitfalls	David et. al. 2024	Comprehensive analysis, highlights AI's impact, enhances diagnostics, personalizes treatment, and improves efficiency.	Challenges include data quality, bias, interpretability, and regulatory issues; they emphasize the need for responsible implementation.
[25]	A robust deep neural network framework for the detection of diabetes	Shahin et al., 2023	It is robust to outliers	A lot of training data is needed
[26]	Internet of Things enabled open source assisted real-time blood glucose monitoring framework	Abubeker et.al., 2024	Non-invasive, IoT-enabled, high accuracy, improves comfort, real-time monitoring.	Limited validation, dependency on technology, potential cost barrier.

capturing the maximum variance present in the dataset. The transformative power of PCA lies in its ability to reduce dimensionality while retaining crucial information, which helps improve computational efficiency. This method plays a crucial role in mitigating the impact of the "curse of dimensionality," contributing to more streamlined and effective machine-learning processes. In essence, PCA serves as a powerful tool to extract meaningful features, allowing for a more concise and informative representation of the underlying data. PCA works by transforming the original dataset into a new coordinate system where the axes correspond to the principal components. The goal is to find the directions along which the data varies the most. Here is a step-by-step explanation of how PCA works.

- Centering the Data: This is achieved by calculating the mean of each feature and then subtracting the mean from each data point, centering the data around the origin.

- Computing the Covariance Matrix: Create a covariance matrix S to capture the relationships between different features as shown in the equation 1.

$$S = \frac{1}{n-1} X_{centered}^{T} \times X_{centered} \quad (1)$$

  Where $X_{centered}$ is the centered data matrix.

- Eigenvalue Decomposition: Decompose the covariance matrix S into its eigenvectors V and eigenvalues D as in equation 2

$$S = V \times D \times V^{T} \quad (2)$$

  Where V represents the directions of maximum variance and D indicates the magnitude of variance along those directions.

- Selecting Principal Components: This is done by ordering the eigenvectors by their corresponding eigenvalues in descending order. Choose the top k eigenvectors to form the matrix V_k, where k is the desired number of principal components.

- Transforming the Data: Multiplying the centered data matrix $X_{centered}$ by V_k to obtain the transformed data $X_{transformed}$ as shown in equation 3

$$X_{transformed} = X_{centered} \times V_k \quad (3)$$

  where each row in $X_{transformed}$ corresponds to a data point in the new feature space defined by the principal components.

The result is a transformed dataset where the original features are replaced by a reduced set of principal components. This new representation retains most of the variability in the data, allowing for dimensionality reduction and simplified analysis while preserving essential information. PCA is particularly useful for visualizing high-dimensional data and enhancing computational efficiency in machine-learning applications.

C. Feature Selection

This study presents a three-step hybrid FS technique, as shown in Figure 2.

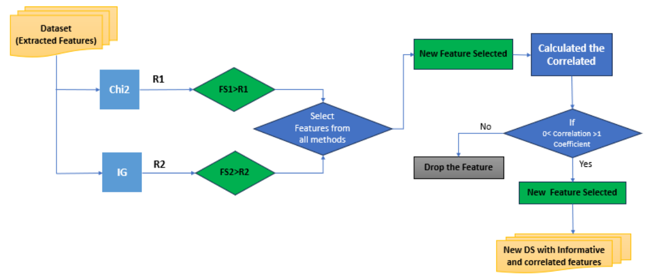

Figure 2. Hybrid feature selection model

To determine the ranking of every feature in the dataset, we begin by employing two FS algorithms: chi-square [28] and information gain [29]. These techniques were selected due to their low propensity for overfitting, adaptability to massive datasets, and quick computing. However, since none of the algorithms rely on a classification system, they are all limited in that they evaluate every feature separately, ignoring any feature relationships. We calculate the average rank for each ranking method (R1, R2) and choose feature sets (FS1, FS2) whose ranked ratings are higher than the mean. Only the characteristics identified in every FS process are taken into account in the next stage. We examine whether these characteristics show commonality or have an association in the last stage. To improve the accuracy of classification, we select only the associated features without redundancy. The suggested FS flow diagram is shown in Figure 3.

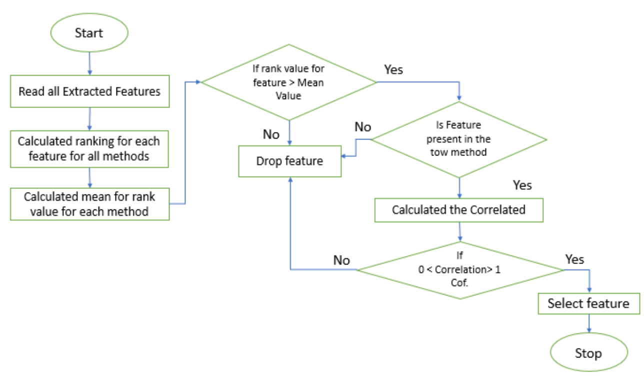

Figure 3. Proposed FS Flowchart

The strength of association among linearly associated features is measured in the study using the Pearson correlation [30], which is just one of the correlation techniques used. The Pearson correlation coefficient (r) is calculated using equation 4. Interestingly, only characteristics with positive values are maintained, while those with negative values are eliminated.

$$P_{r(x,y)} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \tag{4}$$

where $P_{r(x,y)}$ = Pearson correlation coefficient between x and y. And n = number of observations; x_i = the value of x (for ith observation); y_i= value of y (for the ith observation).

D. *Proposed Classification Algorithm*

Random Forest (RF) can be characterized as a compilation of tree-type classifiers, particularly useful in handling datasets with multidimensional features containing many irrelevant variables that can degrade classifier performance. Feature selection becomes crucial for enhancing classifier success, and the RF algorithm addresses this by employing simple probability to select robust features for its inputs. Formulated by Breiman in 2001, the RF algorithm constructs multiple decision trees using subsets of sample data and maps random samples of feature subspaces. Additionally, XGBoost, a high-performance boosting technique, optimizes the loss function through various arrangements and iteratively adds models to a community. This gradient-boosting method focuses on challenging instances for the model, enhancing prediction accuracy. In this study, data classification utilizes an ensemble learning approach with XGBoost and Random Forest as base models, and the final classification is determined through a hard voting mechanism. The parameters, such as kernel, activation function, etc., were tuned to optimize the model's performance. Figure 4 shows the architecture of the proposed classifier model.

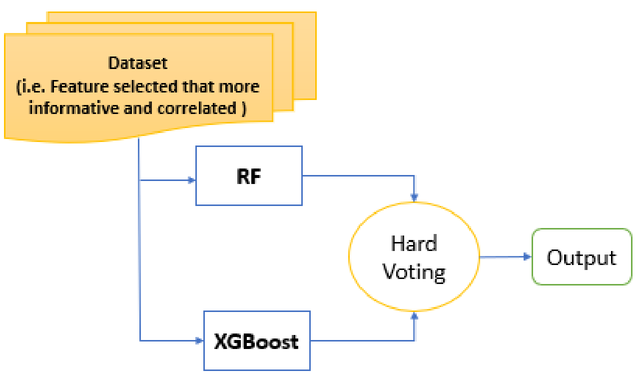

Figure 4. ENS_XGRF Proposed classifier model.

In summary, the integration of RF and XGBoost in an ensemble learning framework, finalized through a hard voting mechanism, offers a compelling approach to data classification. Taking advantage of the various strengths of RF and XGBoost, the ensemble system improves generalization, accuracy, and robustness to various data patterns. The combination mitigates individual model weaknesses, reduces sensitivity to hyperparameters, and provides increased confidence through majority voting. This synergistic ensemble not only captures a wider range of features in the data, but also excels at handling noise and outliers. Overall, the collaborative power of RF and XGBoost, coupled with the simplicity and effectiveness of hard voting, results in a robust and reliable solution for making accurate predictions in diverse machine-learning scenarios.

E. *Particle Swarm Optimization (PASWOP)*

Particle Swarm Optimization (PASWOP) draws inspiration from the collective behavior observed in fish schooling and bird flocking. In this optimization technique, a community of particles is created in a multidimensional space, where each particle's current location corresponds to the expenses that need to be minimized for optimal results. Following every iteration, every item's location and velocity are modified according to a weighted combination of its present velocity, distance from its greatest-known position, and distance from the most optimal position that any particle in the swarming has ever reached globally. In the context of a multivariate solution space, the position and mobility of an object are often denoted as matrices with the symbols p and m, respectively. The $d \times 1$ vectors $(p_{i1}, p_{i2}, \ldots, p_{id})$ and $(m_{i1}, m_{i2}, \ldots, m_{id})$, signify the position and mobility of a particle in a d-dimensional space. Each particle keeps track of its best-known position, denoted as another vector $(pbest_{i1}, pbest_{i2}, \ldots, pbest_{id})$. The location of the most favorable global location amongst all particles is shown in the dth degree by as $gbest_d$. The velocity and position update formulas for a particle in dimension d during the (k + 1)th iteration are determined based on the performance of the kth iteration. These formulas are shown in equations 5 and 6 govern the iterative movement of particles toward optimal solutions within the solution space.

$$\begin{aligned} m_{id}^{k+1} = w \times m_{id}^{k} + c_1 \times rand \times (pbest_{id} - p_{id}^{k}) + \\ c_2 \times rand() \times (gbest_{id} - p_{id}^{k}) \end{aligned} \tag{5}$$

$$p_{id}^{k+1} = m_{id}^{k} + p_{id}^{k+1}, i \in N_p, d \in D \tag{6}$$

The coefficients c_1 and c_2 act as acceleration factors in the multidimensional discovery issue, while D denotes depth and NP denotes the number of participants. The proportionate randomized weight of the departure from the particles' greatest individual achievement and their best aggregate efficiency in the dth dimension is influenced by these parameters. During the search process, the suggested system uses PASWOP with configurable inertia weight, w, to achieve equilibrium in global and local inquiries. Equation 7 is used throughout this study to calculate the state of inertia w, which allows periodic modifications that affect the overall performance of the PASWOP algorithm throughout the optimization stage.

$$w = w_{max} - \frac{w_{max} - w_{min}}{iter_{max}} \times ite \tag{7}$$

In this context, $iter_{max}$ represents the maximum number of iterations and $iter$ is the present number of iterations. The process begins with a large value of w_{max} allowing for an aggressive global search to explore potential good solutions. As iterations progress, w is progressively decreased to fine-tune the search locally, moving closer to the minimal point in the solution space. This dynamic adjustment of the inertia weight facilitates an effective balance between global exploration and local exploitation during the optimization process.

4. Study Result

In this experiment, we use samples of information collected to thoroughly verify our suggested ensemble technique for the diabetic patient tracking system employing the programming language Python. We compare the effectiveness of the proposed algorithm versus many current machine learning (ML) methods, such as Decision Tree (DT) [20], Support Vector Machine (SVM) [20], and Sequential Minimal Optimization (SMO) [19]. A wide range of measures are included in the assessments, such as the F1 score, recall, specificity, precision, sensitivity, precision, and false positives (FP). The objective of this study is to evaluate the efficacy of the methods used and to determine an extremely powerful forecasting method. The confusion matrix is used to do a thorough investigation from which the assessment measures are obtained. A detailed summary of the outcomes is shown in figure 5, which also includes examples of successfully and incorrectly categorized information.

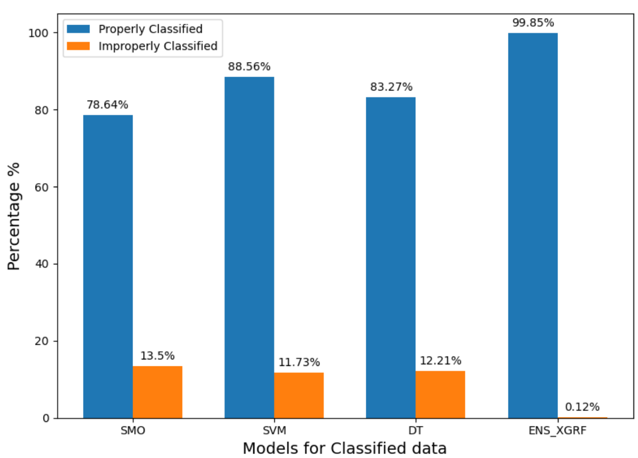

Figure 5. Performance on Properly and Improperly Classified Data

Figure 6 illustrates the comparative learning timeframes of both established and proposed methods. The learning period, representing the time required for algorithms to learn from a dataset, serves as a crucial metric for assessing efficiency. In this context, the suggested model demonstrates significantly quicker training time, clocking in at ”0.019s”, compared to the respective training times of ”0.032s” for SMO, ”0.027s” for SVM, and ”0.051s” for DT. The notably higher training time of decision trees (DT) compared to other machine learning algorithms can be attributed to several factors inherent to the DT algorithm. Firstly, during training, DTs engage in an exhaustive search process, exploring a vast search space to identify optimal decision boundaries for class separation. As the tree grows deeper or incorporates more leaves, the computational burden increases significantly as it evaluates additional decision paths and splits. Moreover, the sensitivity of DTs to dataset features necessitates thorough feature engineering or preprocessing steps, further extending training time. Additionally, tuning DT hyperparameters, such as maximum tree depth or minimum samples per split, requires training multiple trees with different configurations, contributing to prolonged training times. Importantly, the efficiency of the proposed approach, as highlighted in Figure 6, underscores its significance in addressing the demand for rapid dataset training, particularly in real-world applications where time is of the essence.

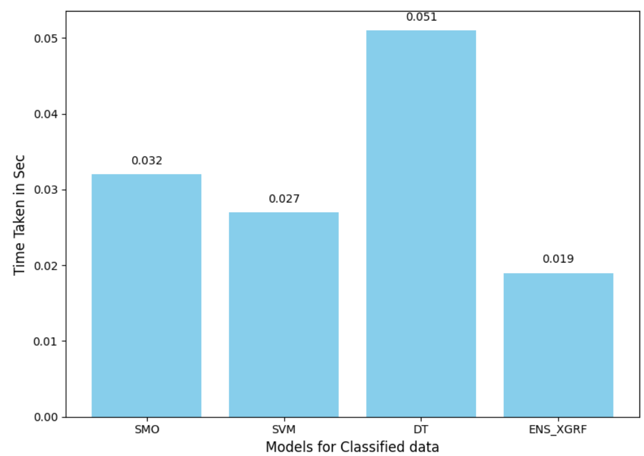

Figure 6. Time Taken by Each Method

Moreover, the proposed method demonstrates superior classification accuracy by correctly classifying a higher proportion of data when compared with the existing methods. The lower rate of improper classifications in the proposed method further solidifies its efficacy in achieving more accurate results compared to the benchmarks set by the existing methods. TableII presents the results of the successful and incorrect categorization of information, as well as the duration of training for the algorithms used in previous studies and those that are being suggested.

TABLE II. Correct and incorrect categorization of data

Strategies	ML-SMO	ML-SVM	ML-DT	Proposed
Correct-classified	78.64%	88.56%	83.27%	99.85%
incorrect-classified	13.50%	11.73%	12.21%	0.12%
Taking time	0.032s	0.027s	0.051s	0.019s

Similarly, Table III presents the results of the comparison of several measures that compare the current and suggested approaches. The proportion of the experiments that meet the results that are correctly anticipated by the recommended approach is used to indicate the technique's performance.

TABLE III. Proposed model with and without FS method

Methods	SMO	SVM	DT	Proposed
TRPO	95.88%	96.81%	94.55%	99.80%
FAPO	2.82%	1.21%	1.12%	0.40%
Accuracy	96.35%	96.58%	94.52%	99.68%
Precision	92.89%	94.35%	96.21%	98.58%
Sensitivity	94.24%	96.36%	95.98%	98.12%
Specificity	95.58%	95.99%	94.84%	98.11%
recall	96.66%	94.88%	95.36%	99.55%
F1-score	93.77%	97.85%	93.55%	98.44%

Equation 8 is used to compute accuracy (ACC), which is a gauge for the number of specimens successfully classified along with the level of match between the input information and the end findings.

$$Acc = \frac{TRPO + TRNE}{TRPO + TRNE + FAPO + FANE} \quad (8)$$

Where (TRPO) is the True Positive value, (TRNE) is the True Negative value, (FAPO) is the False Positive value, and (FANE) is the False Negative value.

Precision (P), which can be found in Equation 9 and is defined as the ratio of correctly categorized incidents to all occurrences of accurately positive data, serves as one of the most significant metrics for accuracy. Precision assesses how well the proposed strategy separates real from false positives through an evaluation of the number of real outcomes compared to the number of expected outcomes. Furthermore, precision analyzes the accuracy of the proposed method.

$$P = \frac{TRPO}{TRPO + FAPO} \quad (9)$$

The ability of the suggested model to recognize any notable element in a dataset is called sensitivity (S). Equation 10 illustrates methods to calculate it analytically by dividing the proportion of true positives (TRPO) by the total of true positives and false negatives (FANE). Sensitivity provides insights into the model's effectiveness in correctly identifying positive instances within the dataset.

$$S = \frac{TRPO}{TRPO + FANE} \quad (10)$$

Recall (RC) of the suggested model is the measure of its capacity to recognize each significant item in a set of data. Equation 11 illustrates how it is mathematically determined as the fraction of TRPO divided by the total number of TRPO and FANE. RC, also known as sensitivity, provides an assessment of the model's capability to correctly identify positive instances within the dataset.

$$Recall = \frac{TRPO}{TRPO + FANE} \quad (11)$$

The f1-score, as represented by Equation 12, combines both precision and recall into a single metric by calculating their harmonic mean. Precision, recall, and f1 score collectively provide a comprehensive evaluation of the model performance. Conversely, it quantifies the probability of a negative result if a negative finding materializes. It is also known as the true negative rate and is a crucial indicator of the model's ability to correctly identify negative instances within the dataset.

$$F1\_score = \frac{(precision) \times (recall) \times 2}{precision + recall} \quad (12)$$

The specificity of the proposed model, as defined by Equation 13, represents the proportion of TRNE to the sum of TRNE and FAPO. This metric is essential for assessing the model's ability to accurately identify instances that truly belong to the negative class within the dataset.

$$Specificity = \frac{TRNE}{TRNE + FAPO} \quad (13)$$

The comparative results from Table III showcase the superior performance of the proposed model on various evaluation criteria. Furthermore, we assess our proposed model with and without the suggested FS method to gauge the impact of this feature selection approach on the framework. The results indicate a notable enhancement in the model's performance, showing an approximately 20% increase when incorporating the feature selection method. This underscores the efficacy and positive impact of the proposed feature selection method in improving the overall performance of the model. Figure 7 and table IV show comparison results with and without the use of the feature selection method in our proposed framework.

TABLE IV. Proposed model with and without FS method

Methods	Proposed with FS	Proposed Without FS
Proper classified data	99.85%	81.96%
Improper classified data	0.12%	8.57%
Taking time	0.019s	0.568s

Ensuring the security of life (SOL) by delivering precise patient information to the hospital is imperative for safe-

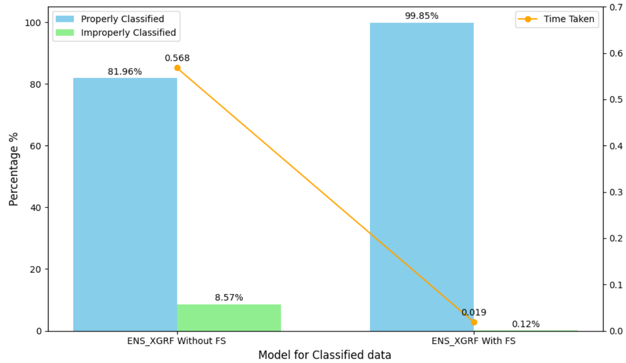

Figure 7. Performance and Time Taken by the proposed ENS_XGRF with and without Feature Selection

guarding patient well-being. Any lapses in compliance can pose a threat to a patient's health. Figure 8 illustrates the outcomes related to the security of life, with SMO achieving 82%, SVM reaching 89%, DT scoring 71%, and proposed 94%. This visualization underscores the effectiveness of the proposed approach, highlighting its superior performance in ensuring the security of patient information.

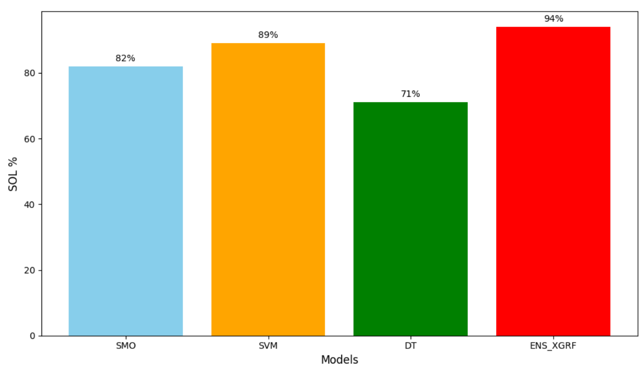

Figure 8. SOL Analysis

Lastly, we computed the AUC-ROC and Receiver Operating Characteristic (ROC) Curve. A binary classifier's ability to discriminate between two classes is visually represented by the ROC curve, which plots the True Positive Rate (Sensitivity) versus the False Positive Rate (1 - Specificity) at various decision thresholds. Condensing the ROC curve into a single statistic, the Area Under the ROC Curve (AUC-ROC) summarizes the classifier's overall performance and potential thresholds. Higher numbers indicate greater selective ability; the value ranges from 0 to 1. A model that achieves an AUC-ROC of 0.5, which is a diagonal line that runs from the bottom-left to the top-right of the ROC space, indicates performance that is no better than random guessing. On the other hand, a perfect model produces a curve that reaches an AUC-ROC of 1. Figure 9 shows the AUC-ROC scores for different models, including the proposed model. The AUC-ROC scores for each model are as follows: Decision Tree (DT) = 0.5, SMO = 0.79, SVM = 0.82, and the proposed model = 0.90. This comparison provides insights into the relative performance of each model in terms of its ability to discriminate between classes. The higher AUC-ROC score of the proposed model suggests that it exhibits superior classification performance compared to the other models evaluated in the study.

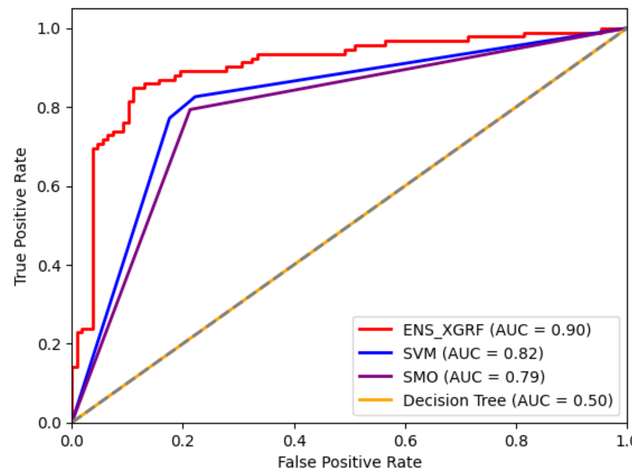

Figure 9. Receiver Operating Characteristic (ROC) Curve for the Proposed Model Compared to Others

5. Conclusion

E-health trackers play a pivotal role in monitoring individual activities and providing essential feedback, especially during critical situations. This article introduces an ensemble learning method for intelligent patient monitoring that allows the assessment of individual dependencies, predicts future health status, and detects potential health declines at an early stage. The normalization approach is used to effectively preprocess the raw dataset effectively. PCA is used to extract features, and chi-square data with information gain are used to calculate each feature's rank in the dataset. Pearson's correlation is used to find the correlation between the features selected by the selection of feature methods. The ensemble learning method that the intelligent system employs for diagnosing uses XGBoost and RF as the base models. The final classification is determined by a hard voting technique and particle swarm optimization (PASWOP). This research shows that the suggested method works better than current methods on several measures, such as accuracy, sensitivity, precision, recall, specificity, F1-Score, TRPO, and FAPO which are represented as follows in percentage terms: "99.68", "98.12", "98.58", "99.55", "98.11", "98.44", "99.80", and "0.40" sequentially. The scalability of the system is emphasized, indicating that it may be extended in the future to include big (i.e. large) and diverse datasets. Researchers studying illnesses in people can benefit greatly from this work, especially in the areas of artificial intelligence and automated predictions. Continuous user input is crucial for future development, ensuring the application remains patient-focused by addressing user needs, refining existing features, and introducing new ones. The emphasis on user privacy is paramount to preventing potential data breaches or leaks, reinforcing our commitment to safeguarding consumer privacy. The study introduces an ensemble learning method for intelligent patient monitoring, aiming to assess individual dependencies, predict future health status, and detect potential health declines early. While the normalization approach effectively preprocesses the raw dataset, challenges in model interpretability arise due to the complexity of ensemble learning models. Additionally, ensuring robust data privacy measures poses a challenge in real-world implementations, particularly in healthcare settings with stringent regulations. Although the scalability of the system is emphasized, extending it to include large and diverse datasets may require additional computational resources and infrastructure. Future research directions include conducting external validation studies to assess the generalizability and robustness of the monitoring system across different patient populations and clinical settings. Enhancing model interpretability is essential to enable healthcare professionals to understand the rationale behind predictions and decisions made by the system. Feasibility studies are needed to evaluate the implementation of the monitoring system in real-world clinical environments, considering factors such as data privacy, regulatory compliance, and integration with existing healthcare systems. Longitudinal studies should be conducted to evaluate the long-term effectiveness and impact of the monitoring system on patient outcomes, healthcare utilization, and overall quality of care. Engaging with healthcare professionals and patients in the iterative design and development process is crucial to ensure that the monitoring system meets user needs, preferences, and expectations while maintaining usability and effectiveness. Ethical assessments are necessary to identify and mitigate potential risks associated with data privacy, bias, discrimination, and unintended consequences of implementing the monitoring system in clinical practice.

References

[1] F. Alrowais, H. G. Mohamed, F. N. Al-Wesabi, M. Al Duhayyim, A. M. Hilal, and A. Motwakel, "Cyber attack detection in healthcare data using cyber-physical system with optimized algorithm," *Computers and Electrical Engineering*, vol. 108, p. 108636, 2023.

[2] A. Ahmed, S. Aziz, U. Qidwai, A. Abd-Alrazaq, and J. Sheikh, "Performance of artificial intelligence models in estimating blood glucose level among diabetic patients using non-invasive wearable device data," in *Computer Methods and Progr in Biomedicine Update, 3, p. 100094*. AMS, 2023.

[3] F. I. Ali, T. E. Ali, and Z. T. Al-dahan, "Private backend server software-based telehealthcare tracking and monitoring system," *International Journal of Online & Biomedical Engineering*, vol. 19, p. 1, 2023.

[4] F. I. Ali, T. E. Ali, and A. H. Hamad, "October. telemedicine framework in covid-19 pandemic," in *2022 International Conference on Engineering and Emerging Technologies (ICEET)*. IEEE, 2022, pp. 1–8.

[5] C. L. Gardner, S. J. Raps, and L. Kasuske, "Cross-sectional analysis of health behavior tracking, perceived health, fitness, and health literacy among active-duty air force personnel," *CIN: Computers, Informatics, Nursing*, pp. 10–1097, 2023.

[6] L. Ji, L. Guo, J. Zhang, Y. Li, and Z. Chen, "Multicenter evaluation study comparing a new factory-calibrated real-time continuous glucose monitoring system to existing flash glucose monitoring system," *Journal of Diabetes Science and Technology*, vol. 17, no. 1, pp. 208–213, 2023.

[7] A. Dhulfiqar and N. Pataki, "April. mec–applications deployment and tcp testing using simu5g," in *2023 International Conference on Software and System Engineering (ICoSSE)*. IEEE, 2023, pp. 38–43.

[8] N. Gharaei, Y. D. A. Otaibi, S. J. Malebary, and A. O. Almagrabi, "A storage optimization and energy efficiency-based edge-enabled companion-side ehealth monitoring system for iot-based smart hospitals," *IEEE Internet of Things Journal*, 2023.

[9] M. A. Azim, E. Merry, J. Gyalmo, and Z. Alom, "Iot in e-health, assisted living, and e-wellness," *In The Internet of Medical Things (IoMT) and Telemedicine Frameworks and Applications (pp.*, pp. 17–38, 2023.

[10] S. Harnal, A. Jain, A. S. Rathore, V. Baggan, G. Kaur, and R. Bala, "March. comparative approach for early diabetes detection with machine learning," in *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*. IEEE, 2023, pp. 1–6.

[11] B. M. Reddy, "Amalgamation of internet of things and machine learning for smart healthcare applications–a review," *Int. J Comp. Eng. Sci. Res*, vol. 5, pp. 0, pp. 8–36, 2023.

[12] A. Motwani, P. K. Shukla, and M. Pawar, "Novel framework based on deep learning and cloud analytics for smart patient monitoring and recommendation," *SPMR). Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 5, pp. 5565–5580, 2023.

[13] S. K. Sharma, A. T. Zamani, A. Abdelsalam, D. Muduli, A. A. Alabrah, N. Parveen, and S. M. Alanazi, "A diabetes monitoring system and health-medical service composition model in cloud environment," *IEEE Access*, vol. 11, pp. 32 804–32 819, 2023.

[14] B. G. Mohammed and D. S. Hasan, "Smart healthcare monitoring system using iot," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 17(01), pp. 141–152, 2023.

[15] H. Pandey and S. Prabha, "February. smart health monitoring system using iot and machine learning techniques," in *2020 sixth international conference on biosignals*. and instrumentation (ICBSII) . IEEE: images, 2020, pp. 1–4.

[16] J. Chaki, S. T. Ganesh, S. K. Cidham, and S. A. Theertan, "Machine learning and artificial intelligence based diabetes mellitus detection and self-management: A systematic review," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 3204–3225, 2022.

[17] P. Rajasekaran and M. Duraipandian, "Secure cloud storage for iot based distributed healthcare environment using blockchain orchestrated and deep learning model," *Journal of Intelligent & Fuzzy Systems, (Preprint)*, pp. 1–16.

[18] S. Dohare, L. Pamulaparthy, S. Abdufattokhov, J. V. Naga Ramesh, Y. A. B. El-Ebiary, E. Thenmozhi *et al.*, "Enhancing diabetes management: A hybrid adaptive machine learning approach for intelligent patient monitoring in e-health systems." *International*

Journal of Advanced Computer Science & Applications, vol. 15, no. 1, 2024.

[19] A. Rghioui, J. Lloret, S. Sendra, and A. Oumnad, "September. a smart architecture for diabetic patient monitoring using machine learning algorithms," *In Healthcare (Vol.*, vol. 8, no. 3, 2020.

[20] V. Deepa and K. Rajeswari, "Analysis on e healthcare monitoring system with iot and big patient data," *Int. J*, vol. 5, pp. 97–102, 2021.

[21] S. Chaturvedi, "Clinical prediction on ml based internet of things for e-health care system," *International Journal of Data Informatics and Intelligent Computing*, vol. 2, no. 3, pp. 29–37, 2023.

[22] P. P. Morita, K. S. Sahu, and A. Oetomo, "Health monitoring using smart home technologies: Scoping review," *JMIR mHealth and uHealth*, vol. 11, p.e37347, 2023.

[23] K. A. M. Nabil, M. A. Islam, A. Al Noman, and M. M. Khan, "March. development of a smart non-invasive glucose monitoring system with spo2 and bpm for diabetic patient," in *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC) (pp.* 0. IEEE, 2023, pp. 193–0197.

[24] D. B. Olawade, A. C. David-Olawade, O. Z. Wada, A. J. Asaolu, T. Adereni, and J. Ling, "Artificial intelligence in healthcare delivery: Prospects and pitfalls," *Journal of Medicine, Surgery, and Public Health*, p. 100108, 2024.

[25] O. R. Shahin, H. H. Alshammari, A. A. Alzahrani, H. Alkhiri, and A. I. Taloba, "A robust deep neural network framework for the detection of diabetes," *Alexandria Engineering Journal*, vol. 74, pp. 715–724, 2023.

[26] A. K. M, R. Krishnamoorthy, S. Gogula, B. S, S. Muthu, G. Chellamuthu, and K. Subramaniam, "Internet of things enabled open source assisted real-time blood glucose monitoring framework," *Scientific Reports*, vol. 14, no. 1, p. 6151, 2024.

[27] A. Rashid, "Diabetes dataset," *Mendeley Data*, vol. V1, 2020.

[28] A. Sikri, N. P. Singh, and S. Dalal, "Chi-square method of feature selection: Impact of pre-processing of data," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 3s, pp. 241–248, 2023.

[29] W. Shu, Z. Yan, J. Yu, and W. Qian, "Information gain-based semi-supervised feature selection for hybrid data," *Applied Intelligence*, vol. 53, no. 6, pp. 7310–7325, 2023.

[30] T. P. Teng and W. J. Chen, "Using pearson correlation coefficient as a performance indicator in the compensation algorithm of asynchronous temperature-humidity sensor pair," *Case Studies in Thermal Engineering*, vol. 53, p. 103924, 2024.

Tariq Emad Ali received the B.Sc. and M.Sc. in Electronics and Communication Engineering from the college of Engineering, Baghdad University. He is a lecturer at Baghdad University, Al-Khwarizmi College of Engineering, Information and Communication Engineering Department. He has 11 published scientific & technical papers, including IEEE Explorer. Mr. Tareq Emad has 12 years of academic, practical and consulting experience in the networking & communication sector. He has experience working with Cisco, Mikrotik, and Huawei networking devices. He currently teaches & conducts research programs in the areas of software computer networks, network automation using Python and Linux, soft computing, cloud computing, intelligent agents, AdHoc networks, wireless sensor networks, routing protocols and security of VANETs, smart antennas in MANETs and WiMAX networks, SDN networks, IoT, IOE, IOV, artificial intelligence, P4 language, and others.

Faten Imad Ali holds a M.Sc. in Biomedical Engineering from the College of Engineering, Al-Nahrain University, Iraq in 2019. She also received her B.Sc. in Biomedical Engineering from the College of Engineering, Al-Nahrain University, Iraq, in 2010. She is a lecturer at Al-Nahrain University, College of Engineering, Department of Biomedical Engineering. She has published scientific & technical papers including the American Journal of Biomedical Sciences, IOP Conference Series: Materials Science and Engineering and the International Journal of Online & Biomedical Engineering. She has 14 years of academic & practical and consulting experience in biomedical research. She currently teaches & conducts research programs in the areas of medical equipment, Biomedical signals, Medical Imaging, Medical Instruments, Three-Dimensional Laser Imaging, machine learning, Optics Electrical Engineering, Biomedical Sensors, Physics Laser, Wireless Sensor Networks, Electrical Circuits, and others. (email: fatenemadali@gmail.com).

Ameer Hussein Morad He received a Ph.D. degree (1998) in Signals and Information Processing from the Institute of Information Science at Jiao Tong University in Beijing, China. He is interested in image and video processing, pattern recognition, computer vision, neural networks, and deep learning. He works as a professor at the Technical Engineering College, Gilgamesh University, Iraq.

Mohammed A. Abdala He received a Ph.D. degree (1998) in Signals and Information Processing from the Institute of Information Science at Beijing Jiao Tong University, China. He is interested in image and video processing, pattern recognition, computer vision, neural networks, and deep learning. Currently, he works as a professor at the Technical Engineering College, Gilgamesh University, Iraq.

Alwahab Dhulfiqar Zoltan He holds the position of Assistant Professor within the Faculty of Informatics at Eötvös Loránd University (ELTE). In addition, his qualifications include the prestigious CCNA certification, as well as IC3 and indefinite proctor/instructor certifications from Certiport. he is accredited as an instructor in a multitude of areas, including CCNA, Linux, Python, CCNP (ENARSI + ENCOR), DEVASC/DevNet, CyberOps Associate, and Network Security, all of which are recognized by Cisco. Furthermore, he is an instructor at the ELTE Cisco Academy, where he imparts his expertise. Additionally, he was honored to hold accreditation as an instructor for Amazon Web Services (AWS), showcasing my comprehensive understanding of cloud/Edge computing solutions.