Machine Learning-Driven IoT System for Patient Monitoring

Adham Mohsen Saeed [1], Yasamin Hamza Alagrash [2], Meriam Noureddine Jemel [3] , Lamjed Ben Said [3]

[1] *National School of Sciences and Information, Manouba University, Bouchoucha, 2000 Le Bardo, Tunis, Tunisia*
[2] *Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, P.O. Box 14022, Iraq*
[3] *LR11ES03 Smart Lab, Computer Science Department, ISG-Tunis, University of Tunis, Bouchoucha City, 2000 Le Bardo, Tunis, Tunisia*

Abstract **- In this paper, a healthcare monitoring system is introduced under IoT based on WiFi enabled Arduino boards to deliver patient data to the cloud through the ThingSpeak platform. It checks abnormalities in health, such as body temperature, level of oxygen, heart rate and vibration, which are checked by the system. A Long Short-Term Memory (LSTM) model is used to generate a synthetic dataset consisting of 14,664 samples, each labeled as either normal or abnormal based on expert assessment. A comparative approach between the traditional machine learning method and the hybrid machine learning model is introduced in this paper. The results demonstrate the effectiveness of the hybrid machine learning model in detecting abnormal patient conditions** in **the traditional machine learning approach, the PCA-SVM (Principal Component Analysis Support Vector Machine) model (achieved an accuracy of 98.87%, while the Random Forest model yielded the highest accuracy at 99%. The integration of hybrid machine learning methods achieved an overall accuracy of approximately 99%.**

Beyond technical performance, this work demonstrates the feasibility of deploying low-cost, scalable IoT systems for continuous patient monitoring in both clinical and home settings. The findings underscore the potential of hybrid machine learning architectures to enhance the reliability and accuracy of automated health diagnostics, contributing to proactive healthcare delivery and supporting the development of more responsive remote patient management systems.

Keywords **- Machine learning, synthetic data generation, Random Forest, PCA-SVM, hybrid machine learning model.**

DOI: 10.18421/TEM153-84
https://doi.org/10.18421/TEM153-84

Corresponding author: Adham M. Saeed,
National School of Sciences and Information, Manouba University, Bouchoucha, 2000 Le Bardo, Tunis, Tunisia
Email: adham.mohsin.saeed@ensi-uma.tn

Received: 08 August 2025.
Revised: 20 January 2026.
Accepted: 05 February 2026.
Published: 27 August 2026.

1. Introduction

Constant population growth poses many challenges for the healthcare sector. A comprehensive literature review identifies limited staffing, inefficient patient flow, extended hospital stays, and inadequate communication techniques as prevalent healthcare challenges [1]. As the urban population continues to grow, smart cities are developing innovative solutions to enhance city life for everyone [2]. A strong healthcare system is crucial for people to work effectively and free from concern, which in turn ensures optimal productivity. Using Internet of Things (IoT) sensors and machine learning techniques, the author [3] conducts policy level study to establish the priorities and strategic aspirations of the participating cities concerning the Smarter Cities Challenge of IBM research. Although it is not health-specific, the study provides the important information on the vision of urban policymakers in regard to the use of AI, data analytics, and smart systems in solving complicated city problems including the delivery of healthcare, emergency response, and equity in the digital health setting. As discussed in the research, numerous cities focused on the public health, safety, and efficiency of the services through AI-powered technology, which demonstrates high demand in terms of data-powered governance [3].

A patient record system plays a central role by integrating appropriate sensing mechanisms to collect both structured and unstructured health data, which can then be effectively analysed using machine learning techniques [4]. For the purpose of communication, smart cities utilize a variety of wireless networks. This category includes wireless LANs, WLANs, proximity wireless, and wireless wide-area networks [5]. The primary components of an IoT-enabled intelligent city healthcare system are highly interconnected. As a whole, “smart” services comprise six main components: Intelligent economies, ecosystems, governments, people, transportation, and lifestyle [6], [7], [8]. This research presents the development of a new real-time health monitoring system that integrates hardware using an Arduino board and applies machine learning algorithms to classify a health status efficiently.

The primary value of this research study is in discovering how an Arduino-based real time health monitoring system can be implemented. Incorporating temperature sensors, heart rates, oxygen levels, and vibration allows for gathering extensive health information. To this end, with the help of the Arduino platform, a cost-effective and highly adaptable hardware system is developed to constantly acquire and communicate sensor information to the central data repository. This system can present a patient’s health’s overall and current status by collecting, not limited to, physiological parameters at a given time. It also serves as the controller that ensures all data taken from the sensors are accurate and collected at regular intervals. Second, the system is also scalable since the combination of hardware and software is designed in modules to fit various medical needs. The collected data were used to train and test machine learning algorithms, demonstrating the proposed system’s effectiveness in supporting data-driven health monitoring. The primary contributions are listed below:

Develop an Arduino-based integrated IoT health monitoring system to minimize the costs of healthcare data collection and enable continuous, real-time monitoring economically efficient tracking of vital signs for proactive diagnostics and remote patient management.

To support robust evaluation, an LSTM model was designed to generate a realistic synthetic sensor dataset, and a logic-based hybrid system integrating multiple machine learning algorithms was developed to improve health status classification. Two ML strategies comparison: Naming tradition and hybrid. To find the accurate model used to diagnose and classify the health status in the real-time mode, ML models, including linear regression, SVM (Support Vector Machine), Random Forest (RF), and SVM-PCA are adopted.

Above all, this study’s primary research question is: *Which machine learning model is most suited for a real-time system that can achieve accurate results for a public health numerical dataset?*

In the paper, a comparative model of classical machine learning techniques in contrast with a logic-based hybrid one is presented and the task is to improve automated health monitoring and decision support.

In the study, emphasis is put on the creation and testing of a series of supervised learning models, that is, Linear Regression, (PCA-SVM), and Random Forest, in both approaches. These models were trained with datasets of patients that possessed information pertinent to diagnosing abnormal states of health as well as the kind of physiological information that is conducive to early alerts in regards to unhealthy states.

This paper is organized as follows: Section II highlights the latest IoT health solutions for therapies and machine learning and identifies research gaps that motivate proposed system. Section III describes the system’s design using Arduino together with sensors coupled with ESP8266 as a remote data transmitter presenting the proposed system. Section IV details models like Linear regression, SVM, PCA-SVM, and Random Forest, by describing their settings and health status classification outcomes. Section V points out the differences between model performance and the specific usage of each algorithm. At the same time, Section VI presents concrete ideas for general applications and real-life usage, including more complex datasets for trials and the expansion of sensor numbers. Section VII revisits the topic of affordable health monitoring as the study’s contribution to healthcare technology using IoT and ML. The structure effectively guides readers from general context research to practical implications.

2. Related Work

Integrating Remote Health Monitoring in smart cities necessitates the use of IoT and machine learning (ML) technologies [9]. These procedures often utilize interconnected IoT devices to conduct necessary predictive and prognostic analyses, remote monitoring, surgeries, and preventive measures.

Remote healthcare monitoring necessitates customized models and designs that are integrated with the patient record system to facilitate effective data management.

Among the forthcoming healthcare solutions that remote health monitoring aspires to develop are apps for glucose level sensing, electrocardiogram (ECG) monitoring, blood pressure (BP) monitoring, temperature (T) monitoring, oxygen saturation (OS) monitoring, rehabilitation systems, medical management, wheelchair management, and more.

Research [10], linked imaging, patient care, medication distribution, hospital procedures, tests, pharmacy control and diagnostic were among the many medical gadgets introduced by the IoT during the COVID-19 pandemic [11]. Along with this came improvements in patient safety through the use of I-patient biological services, ultrasound, X-rays, glucose meters, smart beds, blood gas analyzers, and thermometers [12].

With the help of predictive analytics and ML-based remote healthcare monitoring services, hospitals can help patients recover more quickly and return home sooner. It is possible to simply develop a risk stratification model with the help of predictive analytics, which will enable more resources to be dedicated to care for a group of patients with high-risk indicators [13].

These models utilize both static and dynamic patient data to predict potential outcomes and develop plans to mitigate the impact of any issues. Results from ML-based applications are validated by the statistical data supporting deep learning as an integral component of ML [14], [18].

According to several studies [15], [16], [17], clinical variables and vital signs can be effective in predicting readmission. However, the high-quality dataset is still a challenge. Contemporary facilities use data examination and collection machines to facilitate data collection and share in more advanced information systems. However, noisy data, missing data, data transmission loss, should still be considered [18].

In recent years, remote sensing (RS) and monitoring systems have increasingly incorporated machine learning, deep learning, and big data techniques. Prior studies [19], [20] have focused on explaining both the theoretical foundations and practical applications of Artificial Intelligence, with an emphasis on deep learning and large-scale data handling. Environmental monitoring and control, or RS, is using newly available massive volumes of data to reveal hitherto unimaginable opportunities for global surveillance of both natural and artificial processes. However, this work suffers from complexity, particularly training and testing.

Electronic Health Records (EHRs), clinical text, electronic medical reports (EMRs), medical photos, and social media are all examples of publicly available health-related data that have greatly aided DL's ascent to prominence in the health sector [21]. Discussing the process of machine learning implementation in the medical sphere, [22] provides essential ideas of how ML/DL systems are going to change the patterns of medical data processing and decision support.

They show that it is possible to maintain the diagnosis of the disease, the optimization of Virtual Reality (VR) and the individual course of management by previous data-driven analytics sources of both structured and unstructured sources. The paper profiles the architectures of the most important models (Convolutional neural networks, gradient boosting machine and the hybrid ensembles), but it also provides a broad description of the implementation issues, such as data quality, explainability of the model, and the implementation in the real medical practice. The survey presented by [23] demonstrated that machine learning architectures deployed within multimodal frameworks are capable of integrating heterogeneous data sources, including sensor data, electronic health records, and medical imaging, thereby enhancing clinical decision-making. These systems tend to operate using deep learning and decision trees and ensemble models in order to identify deviations and patterns in health more accurately.

To a similar end, [24] presents a generic ML framework to remotely monitor patients (RPM) and interpolates the whole pipeline that includes data collection, feature extraction and predicting as well as decision-making. By integrating supervised learning and reinforcement learning, the framework monitors vital signs to provide early predictions and clinical alerts. The data's unique properties and variety of goals make the use of ML and signal processing methods possible [25]. The role of data analytics and machine learning in telemedicine system based on wearable devices usingIoT is addressed by [26] with the special emphasis on the importance of the ML algorithms in processing the stream of data transmitted by physiological sensors.

In the paper, they defined the architecture required by the system in topics of real-time analysis with cloud-based platforms and edge devices. It is stated that anomaly detection, pattern recognition and time-series analysis based on ML techniques could be utilized as organizational tools allowing automatic diagnostics and predicting healthcare outcomes. A multimodal ML-based solution for Alzheimer disease prognosis that integrates the data of imaging studies, clinical reports and cognitive test results [27]. The authors use some of the ML algorithms, such as a SVM, random forest, and gradient boosting to enhance the accuracy of Alzheimer stages classification. Their system indicates how ML can successfully handle heterogeneous health data, and how to use them to obtain higher predictive accuracy than any other individual model.

A hybrid ML-based multi-disease detection is proposed in [28], utilizing supervised learning algorithms such as logistic regression, decision trees, and K-nearest neighbors (KNN) as one of the methods used and neural networks. The latter were trained using clinical datasets on several diseases. The research demonstrates that the combination of these models into a hybrid ensemble provides a great improvement of classification performance. Machine learning is the focal feature of pulling and processing trends across illnesses, and over fitting and non-generalization has been considered the most crucial issue.

A remote health monitoring system based on deep learning with the assistance of IoT is developed in the study [29], in which physiological indicators of the heart rate, oxygen saturation, and the body temperature are measured periodically, and subsequently, they are analyzed using the convolution, neural networks with attention mechanism. To achieve this, a deep learning approach may be applied to enable the system learn the spatial and temporal structures in the data, and the stage-1 accuracy of the anomaly detection is currently at the level of ~98%. The reasoning behind their approach is that IORD sensing and deepML are viable when it comes to the creation of scalable, patient monitoring systems working in real-time.

As presented in [30], a hybrid cloud model ML applies XGBoost and Bi-LSTM to predict the existence of heart disease on streams of sensor data at a remote patient care center [30]. The streaming health data is presented to the model with XGBoost model being a steady feature (e.g., demographic records or clinical records) and the Bi-LSTM (Bidirectional Long Short-Term Memory) models are a time series pattern in the current Bio- monitor platform.

This amalgam chooses both available machine learning and deep learning, providing higher precision and versatility of patient tracking in the long term. The system accuracy reads 98.5, which demonstrates that ML may be in the roots of the high-performance diagnostic systems in telehealth.

These studies allow seeing the high diversity of ML algorithms applied to remote healthcare systems studied, between the supervised learning (e.g., SVM, RF, KNN), the deep learning (CNNs, LSTMs), and the hybrid ensemble models.

Machine learning offers auto feature extraction, classification, identifies trends and is able to predict diagnosis. Despite the tremendous success achieved so far, the challenges that remain are that of ensuring data diversity, effective computation, model validation and privacy enforcement within an operating environment.

3. The Proposed System

The current study introduces an IoT-enhanced framework that operates in real-time mode, engaging patients in answering health-related questions and gathering the results to classify patients’ health status accurately. Pulse rate, oxygen saturation, body temperature, and vibration are some physiological parameters monitored using a network of sensors in the system.

The system utilizes readily available, low-cost items, such as Arduino devices, in conjunction with efficient forms of communication, including ZigBee and ESP8266, to minimize power consumption and reduce implementation costs in resource-constrained environments.

The proposed framework offers instant supervision, flexible sensor innovations, and a machine learning approach, including traditional and hybrid methods, that enables the healthcare system to respond rapidly while ensuring the framework remains scalable.

The proposed system is presented in Figure 1, which begins with the collection of various sensor data (accelerometer, gyroscope, temperature, heart rate, and oxygen levels). This data is then aggregated into a healthcare dataset via a cloud platform. This raw data undergoes essential cleaning and preprocessing to make it suitable for analysis. Subsequently, two distinct machine learning approches, a traditional approach and a hybrid system, could involve ensemble methods, which combine multiple traditional models. The ultimate step involves evaluating the results to compare and assess the performance of these two machine learning approaches on the healthcare data.

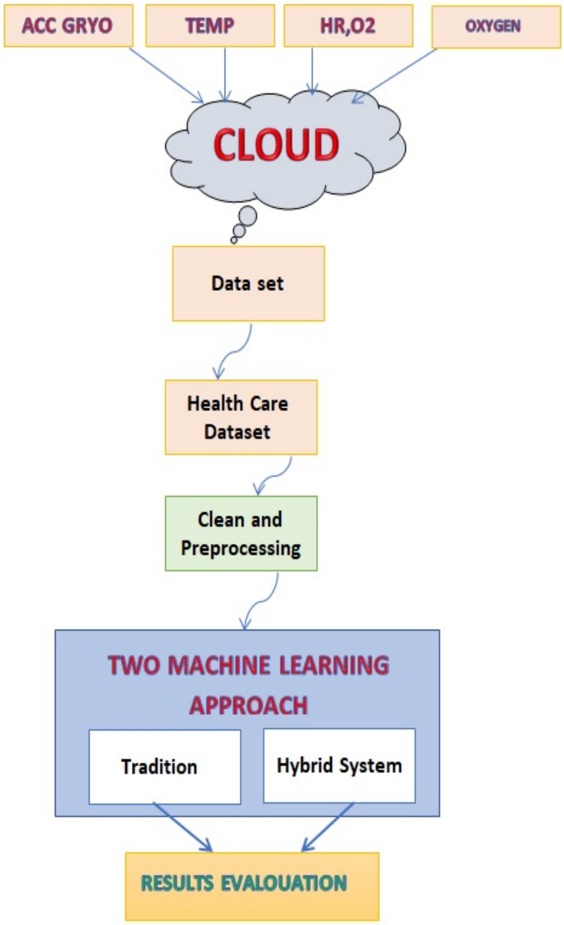

Figure 1. The process of the proposed approach of real-time monitoring of the healthcare system

3.1. Storage Component

This phase begins with a user creating an account on the ThingSpeak IoT platform by using cloud storage, after which sensors gather data and store it. Figure 2 shows the assembled data process. In the Arduino system, vital signs and environmental parameters are sensed using various sensors and modules connected to the Arduino UNO, which is part of the proposed system. It features an LCD that displays real-time data, including temperature, heart rate, and oxygen levels. The interposer features an embedded keyboard for setting up inputs or manually entering commands. The accelerometer and gyroscope module record the user's motion and positioning, as well as features such as walking or a potential fall. A thermistor tracks the body's temperature, which is crucial in monitoring a person's health for certain diseases. The heart rate and oxygen sensor refer to the combination of pulse oximetry used to assess the heart rate; PPG is used for detecting SpO2 among other string stream EOFs. For further data entry into a remote server or cloud, the system is equipped with an ESP8266 Wi-Fi module, making it suitable for IoT systems. The Arduino Uno is the main microcontroller, which only reads data from the sensors and controls the data output to the LCD and the wireless module [31]. All of them are wired through a breadboard, enabling an easy flow of information between the sensors, display, and Arduino.

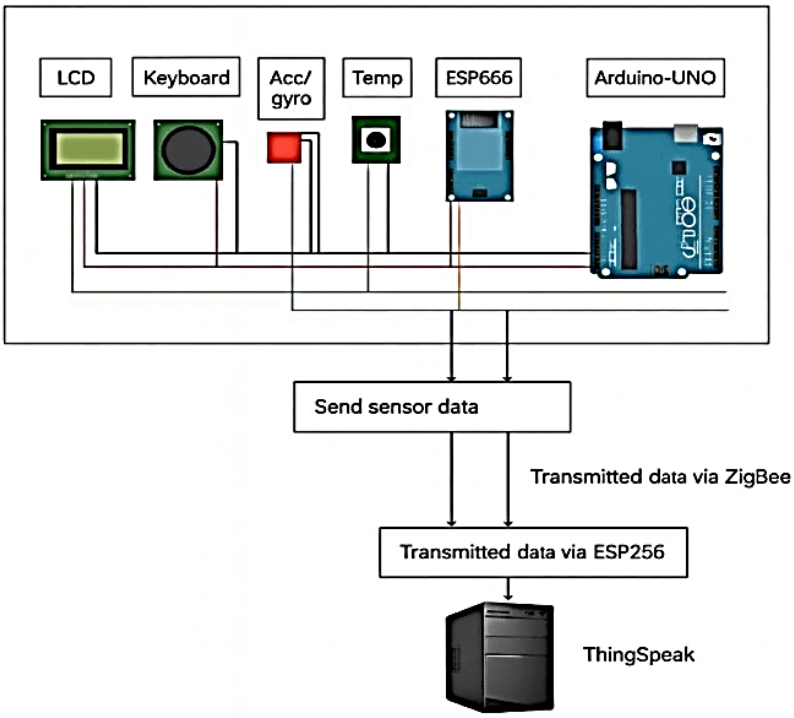

Figure 2. Real time dataset collection component

3.2. The Security Channel between the Collection Hardware and the Cloud

The study employs ThingSpeak, which is an Internet of Things (IoT) analytics system designed by MathWorks, to handle data transmission as well as cloud-based storage in a remote monitoring system. The platform is real-time data acquisition and analytics-friendly, and data is transmitted safely using TLS/SSL encryption, which is an industry-standard method of securing data on the move. It is a security mechanism that blocks unauthorized interception provided in case of communications between the edge device and the cloud. In the case of MQTT based communication, ThingSpeak provides the support of encrypted connectivity using port 8883 and allows the usage of authentication credentials-client ID, username and password, which are securely controlled through MATLAb vault system. This boosts confidentiality especially in medical monitoring cases, where confidentiality of a patient is being relayed constantly.

Furthermore, ThingSpeak gives the creators capability of creating personal channels where it is possible to control the visibility and access on finer scale. This is useful especially in application to healthcare where data integrity and confidentiality is paramount. Access is controlled through API keys, which authorize users for read and write operations on specific data channels [32]. The ThingSpeak platform provides a secure infrastructure suitable for healthcare applications due to its built-in encryption, robust authentication, and support for data-secure communication channels [32]. Its ability to balance strong security with powerful data analytics makes it an ideal choice for research in smart health monitoring systems designed to detect physiological anomalies.

Data transmission security is enforced at the protocol level; interactions via HTTPS or MQTT use TLS encryption to prevent interception or modification [32]. Specificaly for MQTT, ThingSpeak uses port 8883 for MQTT over TLS, ensuring secure data transfer even from resource-constrained IoT devices [32]. Security can be further augmented at the device level through hash-based authentication and AES-256 encryption [32].

3.3. Classification System Component

In this section, the classification methodology is presented starting with the stage of generating the dataset through preparing, and the proposed comparative approach.

3.3.1. Data Preparation

The cleaning and preparation of such samples involve several steps to create the final dataset, facilitating further analysis and modeling. The initial data set consists of 50 samples, each containing several fields: Temperature, DateTime, ID, PatientName, Address, and other non-numeric fields, which are considered non-relevant. The discarded fields of interest include Time, Latency, ID, and Status. In contrast, the fields of interest retained are Heart Rate, Oxygen Level, Temperature, and Vibration, as they yield numerical measures that are practical in health status monitoring.

3.3.2. Deep Learning for Synthetic Data Generation

To address the limited size of the collected data, a deep learning model is utilized to generate a synthetic dataset to expand the collected data for robust system. The synthetic data generation process is detailed in the algorithm which is presented in Figure 3.

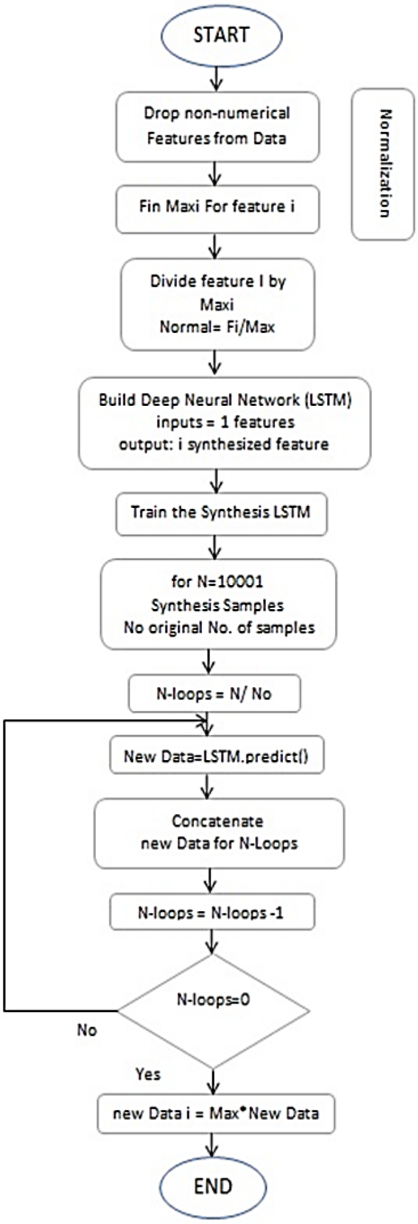

Figure 3. Synthetic data generation flowchart

In order to generate data, a LSTM model architecture presented in Figure 4 was used. This architecture consists of an input layer, 2 consecutive layers of Long Short-Term Memory, a dropout layer to avoid overfitting, and an output generation layer of dense type.

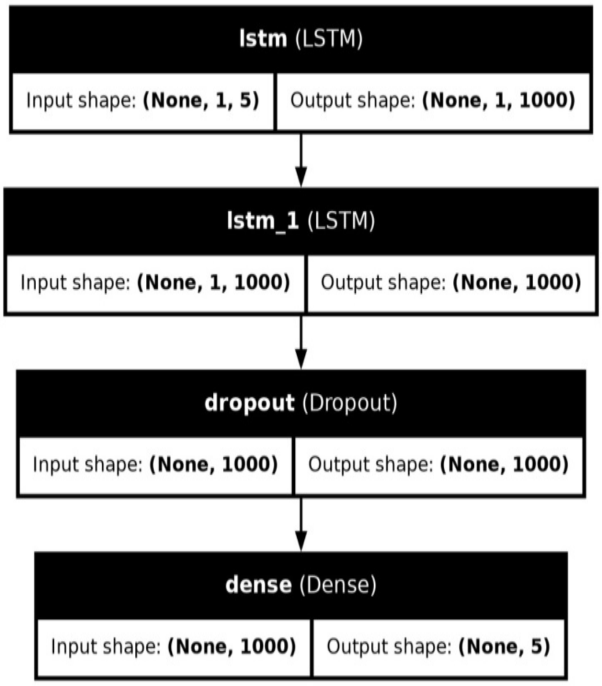

Figure 4. The used model to generate data

The LSTM model is developed to process sequential shaped input data under (None, 1, 5). In this case, None is used to represent the batch size enabling the model to use an arbitrary number of samples simultaneously. The value 1 indicates the number of time steps per input sequence, and the value 5 refers to the number of features at each time step. In other words, each sample consists of a single time step described by five distinct features.

The architecture includes the first LSTM layer that captures temporal dependencies in the input data. It receives input with the shape (None, 1, 5) and produces output with the shape (None, 1, 1000). Since this output corresponds to the same length output (i.e., one time step) with each time step being replaced with a 1000-dimensional feature representation, the implication is that the layer computes such a sequence and does not in any way act differently on time steps of different lengths. The Output dimensionality is related to the number of LSTM units in this layer.

The second LSTM layer is stacked on top of the first one and receives as input the output of the previous layer with shape (None, 1, 1000).

At this point, the resulting tensor has the shape (None, 1000), indicating that the time dimension has been removed. This behavior is configured using the argument return_sequences=False, which ensures that only the final hidden state is passed forward. This setting is particularly suitable when the LSTM is followed by downstream fully connected layers. Whether for classification or regression tasks, this approach is widely used and appropriate for preparing the network for final prediction."

Then there are the repetitive layers with a Dropout layer regularization of the weights that can be implemented to avoid overfitting. It takes in an input of (None, 1000) and proceeds to have the same structure of output. The input units in the course of training are randomly zeroed at every update of the network and this works in enhancing the network generalization and robustness.

The last architecture layer is Dense (Fully connected) which is the last structure of the processed output of Dropout architecture layer. It transforms the Input of shape (None, 1000) to (None, 5). Output dimension of 5, is a number of columns in a classification problem or in multi-output regression problem.

The dataset is labeled by adding an additional column (feature) that represents the label. The labels are defined as follows 0 for standard and 1 for abnormal. Ground truth operations were performed to assign these labels. The normal ranges, established through the ground truth process, are as follows:

- Heart Rate: Between 60 to 100
- Oxygen Level: Between 95% to 100%
- Human body Temperature: It is between 35 and 38
- Vibration: less than 60

This study has 14,664 samples, equally divided between healthy and unhealthy classes. Such separation of the dataset provides balanced training for the machine learning models from both classes, enhancing strong reliability and model generalization.

The system utilizes some sensors to continuously capture data and relay it to an Arduino board, which then inserts the data into a database that serves as a centralized data source for viewing both current and historical data. It can be used in trend analysis, machine learning models for predictive analysis, and remote monitoring of uploaded data, making it valuable for all fields of healthcare, manufacturing, the environment, and research.

3.4. Predictive Analysis Using Machine Learning Approaches

To evaluate the performance of any machine learning model, it is essential to split the dataset into two distinct parts: Training and testing data. The presented dataset, with 14,664 records, is divided at an 80%/20 % ratio, with the data equally divided into two classes (0 represents healthy and 1 represents unhealthy). For the proposed monitoring system to function effectively, it is crucial to accurately classify a person's health status using data acquired from sensors, for which machine learning classifiers are applicable. Physiological parameters, including pulse rate, pulse oximetry, temperature, and vibratory signals, are continuously recorded and analyzed through the system. Due to this, there is a high degree of variability in these parameters, and traditional rule-based approaches are insufficient for modeling the healthy (class 0) and unhealthy (class 1) states.

This study develops a comparative approach between traditional machine learning and a hybrid system for health status classification, leveraging ensemble-based logical integration to enhance predictive accuracy and robustness in real-time patient monitoring.

3.4.1. Traditional Machine Learning

Traditional Machine Learning refers to conventional machine learning algorithms such as Support Vector Machines (PCA-SVM).

These classifiers learn from labeled data and leverage the strengths of their underlying algorithms to produce highly accurate results

Supervised learning algorithms in this system including Linear Regression (LR), Support Vector Machine and Random Forest are classifiers that are trained on the balanced dataset, thereby preventing bias in learning data features and patterns, as well aspredicting the health class of new data points. The classifiers used are selected due to their suitability in managing vast numbers of high-dimensional features related to varied data inputs and in providing dependable classification results. For example, Random Forest and ensemble approaches are apt for high-dimensional data and non-linear data.

3.4.2. Hybrid System

At this stage of the model pipeline, an ensemble hybrid learning approach is applied, in which multiple machine learning models are combined to produce a final decision.

The heterogeneity of the model architectures is exploited to achieve an optimal performance of the encompassing classification by covering a wide range of data distributions and reducing possibilities of biases limited to a specific type of model.

An ensemble brings together a multiplicity of machine learning models where each presents a unique asset:

PCA-SVM: It combines Principal Component Analysis (PCA) that is employed to cut down dimensionality and (SVM) classifier. It is particularly suitable in data manipulation that has got a high dimensionality where optimal margins of noise have negative impacts to analyses.

Logistic Regression: The Logistic Regression is a linear form of the classifier that is employed in the modelling of the probability of categorical ones. It is comprehensible and interpretable and, therefore, suitable in detecting linearity in form of relationships of variables present in data.

RF: A well-known ensemble method, which uses decision trees, and is frequently extolled not only because this method can model complex interactions (which are not linear), but also since the method can rank features. It also has the effect of reducing over fitting through averaging models.

SVM: One of the most sophisticated classifiers which works efficiently within the high-dimensional space and functions ideally when there is an obvious boundary dividing classes.

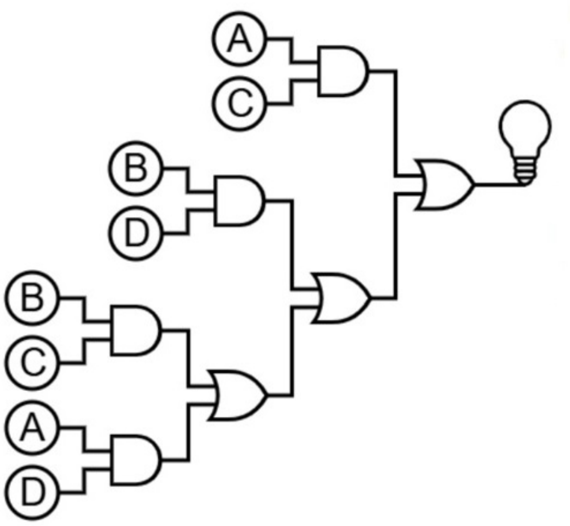

Figure 5. Logic hybrid system approach

The application of those heterogeneous models guarantees the variability in the task of classification by offering several ways of addressing it and letting the hybrid structure learn various patterns with better generalization on the data that is not seen.

Figure 5 indicates the flow of logical construct of this suggested hybrid learning framework, where the process of interaction and decision fusion among the individual classifiers is mentioned.

A = PCA_SVM Model
B = Logistic Regression
C = RF
D = SVM

Using the following Boolean equation:

Logical Model = (A AND C) OR (B AND D) OR (B AND C) OR (A AND D)

In the proposed hybrid architecture, the predictions from individual machine learning models are integrated using Boolean logic gates. This logical framework functions as a meta-classifier, synthesizing the outputs according to predefined rules. The AND operations relate to a consensus mechanism, thus, a sub-expression will be identified as equal to the value true in case, and when there is a coincidence in the categorization of both models that form it. This would make the decisions made more reliable since the confirmation is two sided. Taking an example, say that the output given by the model A (PCA-SVM) and the model C (Random Forest) are to be calculated in detecting a particular pattern and that the same classification has been produced by the two models, then the combined variant A C is to be selected with higher confidence than when the detection is calculated by using either of these two models alone. This kind of process assists in reducing false negative or positive that might be due to the weakness of the personal models.

In contrast, OR operations kindle certain level of flexibility and complementarity to decision-making. They make the model end positively where there is at least satisfaction of one of multiple combinations of models, e.g. A C or B D, is an indication of a decent shot on a target class. It is particularly useful in the case where different combinations of models could be used specializing in detecting distinct patterns of data or subsets of features. Suppose to take a simple case, A C may be a very successful method of ascertaining one kind of abnormality and B D the more successful in determining that which afflicts another kind. The other combinations of B with C and A with are backup, but not alternative channels and this makes the overall system more robust. It operates like a majority based or a weighted ensemble with the implementation being done based on logical expressions.

When one pathway (e.g., A·C) does not result in the proper evaluation, the succeeding pathways, e.g., B·D, B·C, or A·D might still lead to the proper evaluation, thus resulting in better recall and precision.

Such logical grouping of model outputs eventually allows hybrid system to behave as structured ensemble classifier, having the ability to combine the complementary strengths of individual models in a robust and interpretable fashion.

4. Testing and Evaluation

The testbed resembled an electrical substation. IoT environment is created to represent a real-world medical test. Figure 6 and Figure 7 illustrate components of the test setup.

Figure 6. Snapshot of medical testing

Figure 7. Example of a medical test

Access to an online database for real-time communication between the patient-embedded unit and the IoT server. Measured signals from medical sensors are saved in an online database. For the executing prototype, a domain is set up on a host for testing purposes, and a database is created using ThingSpeak.

Through the use of synthetic dataset generation, a total of 14,664 samples were produced overall. The process of rendering the dataset balanced is known as data oversampling, focusing on the unhealthy samples, specifically the minority class in this case. Following the production of the first data set, healthier skin tissue samples (class 0) were recorded in greater numbers than unhealthy ones (class 1), which led to a skewed data set that may skew the learning algorithm in machine learning.

To address this, the number of unhealthy samples was artificially set equal to that of healthy samples to achieve a more balanced classification ratio for both classes. This oversampling technique typically involves duplicating the existing data samples or generating artificial samples for the minority class. Random oversampling involves duplicating the instances, while SMOTE synthesizes new sample points by taking a percentage of the distance between the randomly selected near neighbors. Balancing the data, as the labelled data has no balance, is shown in Figure 8.

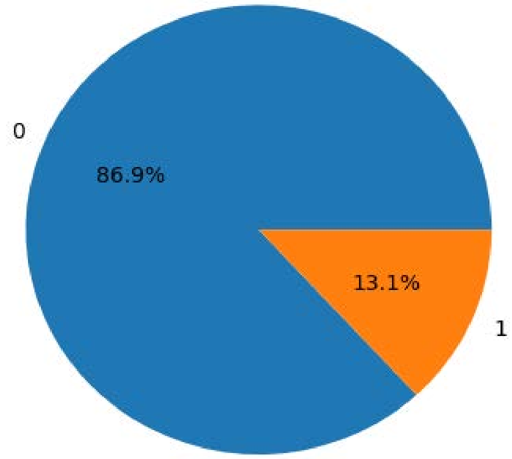

Figure 8. Unbalanced dataset

In order to balance the data, the label class with fewer samples, in this case, the abnormal or 1, is up-sampled to have the same samples as the regular class or label 0. Figure 9 shows the distribution of the 0 and 1 labels after balancing.

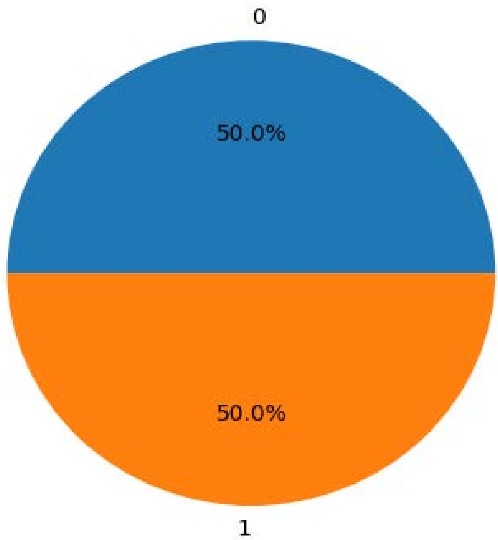

Figure 9. Distribution of the 0 and 1 labels after balancing

In a traditional ML model, the LR model belongs to the linear family of machine learning models, which incorporates the sigmoid function to classify the data. In the sigmoid rule, the data labels are classified as class 0 or class 1. Table 1 shows the results of the LR.

Table 1. LR classifier results

Classifier	LR
Accuracy	91.974%
Precision	93%
Recall	92%
F1-Score	92%
AUC ROC	92.13%

The proposed SVM model achieved an accuracy of 96.96% for classifying the seen data and the unseen data, demonstrating its efficiency in differentiating between transformed sensor data from healthy and unhealthy individuals. The proposed system shows high accuracy due to its ability to recognize the features that extracted from the sensors, including heart rate, oxygen level, temperature, and vibration, yield relevant information which may result in the capability of supporting sound classification. The results obtained after testing the SVM model were analyzed, and Table 2 explains the results of using the SVM classifier.

Table 2. SVM classifier results

Classifier	SVM
Accuracy	96.38%
Precision	96%
Recall	96%
F1-Score	96%
AUC ROC	96.39%

PCA was performed to determine most of the features while significantly decreasing the dimensionality of the data to enrich the feature space for the SVC model. PCA attains this feat by underpinning a criterion of the highest variance, which defines the directions or the feature space. Built on this transformed data, PCA can help detect essential patterns with reduced noise and irrelevant features. However, to enhance the class discrimination in the sensor data, PCA was applied for the healthy and unhealthy classes; similarly, after performing PCA on the features, as well as the SVC. Clemson's enhanced PCA-SVM pipeline analysis demonstrated an increased accuracy of 98.7% on unseen test data. Table 3 indicates that the model performs well.

Table 3. PCA-SVM classifier results

Classifier	PCA-SVM
Accuracy	97.87%
Precision	98%
Recall	98%
F1-Score	98%
AUC ROC	97.94%

For the synthesized dataset, another model was developed using a Random Forest Classifier, which also achieved a notable performance of 99.6% accuracy. Random Forest achieves its goal by building several decision trees during training and then making each by random sampling through a random subset of features and observations. Random Forest, the prediction of which combines the results of these individual trees through averaging or "voting", is less likely to overfit as a single decision tree. This approach yields vast generalization, producing good results with both the training data and new datasets.

The corresponding Table 4 shows no misclassification between the Random Forest model and the actual values tested. This perfect score reveals the efficiency of Random Forest on this dataset and shows that, indeed, the fixed threshold labeling in the data facilitated easy labelling of the sample.

Table 4. RF classifier results

Classifier	RF
Accuracy	98.96%
Precision	97.6%
Recall	98.6%
F1-Score	98.6%
AUC ROC	98.6%

Logical hybrid system is assembled in the same sequence of machine learning model, i.e., LR, Support Vector Machine, PCA-SVM, and Random Forest. This architecture makes use of ensemble-based system of logic inferencing that entails that the outputs of the many base models join through specified logic operations to bring about the final classification decision. It will be seen that the system will make use of the stronger points in each of the models as well as remedying the weaknesses in each of the models. Such rationale combination results in an astonishing classification accuracy that ends up at 99.86%.

Table 5. The report of the hybrid system

Classification report	Precision	Recall	F1-score	Support
False (Healthy)	1.00	1.00	1.00	1441
True (Not Healthy)	1.00	1.00	1.00	1482

AUC ROC Accuracy: 0.99863127

The classification report (Table 5) shows that model performance is of the high standard. Precision, recall and F1-scores are all equal to 1.00 and both classes and the AUC-ROC value is close to 1.0, which means the model is performing close to perfect specifications. Such performance level implies the low level of classification mistakes which proves extremely high accuracy, robustness, and reliability of the model.

These results are also confirmed by the confusion matrix of the logical hybrid system, as is illustrated in Figure 10. Based on the values on the matrix, a number of essential insights on the predictive behavior of the model are to be drawn:

- **High Accuracy:** The model correctly classified a very high number of instances (1439 + 1480 = 2919) out of a total of (1439 + 2 + 2 + 1480 = 2923) instances. This indicates high **overall accuracy**.
- **Low Errors:** Both the False Positives (2) and False Negatives (2) are very low, suggesting that the model makes very few mistakes in both types of misclassifications.
- **Balanced Performance:** The model performs well for both “Healthy” and “Not Healthy” classes, with a low number of misclassifications for both.

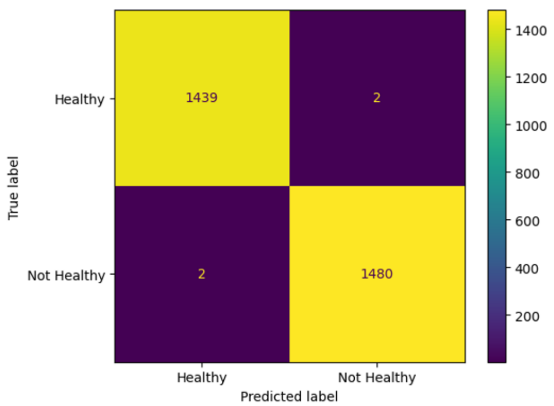

Figure 10. Confusion matrix of the hybrid model

5. Discussion and Analysis

This work presents an efficient embedded system-based real-time healthcare monitoring system that utilizes an Arduino board with four biosensors and Wi-Fi technology. The signals will be entered into a ThingSpeak IoT environment to save patient datasets. The outcome of this study provides insights into determining the optimal machine learning algorithm that operates smoothly in real-time conditions and with numerical data sets.

Figure 11 shows a comparison graph of classification accuracies of various machine learning models such as LR, SVM, Principal Component Analysis with SVM (PCA-SVM), RF and the proposed Logical hybrid system. The figure shows the obvious increase in performance realized by using the hybrid method.

The hybrid system had the highest accuracy of 99.86 as indicated in the chart, as opposed to all individual models. Random Forest was second with 98.6% followed by PCA-SVM which came in at 97.94%. SVM model achieved the highest accuracy of 96.39 by itself, whilst Linear Regression performed the worst in the group at an accuracy of 92.13.

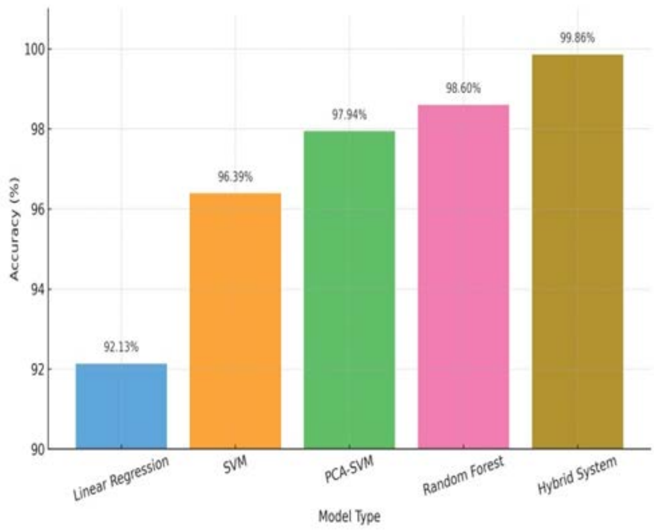

Figure 11. Accuracy comparison of traditional models and the proposed hybrid system

Table 6. Evaluation of proposed solutio

Ref	Methodology of IoT	Approach	ML Method	Dataset	Accuracy
[26]	Static model	Multi disease detection: Heart Disease, Liver Disease, Diabetes Disease	LR	Heart Disease Data (from UCI Repository), Liver Disease Data (from Kaggle Repository) and Diabetes Data (from Kaggle Repository). T	80.76%
			DT		78.57%
			MLP		82.41%
			KNN		85.46%
			Hybrid		97.42%
[27]	Remote model; 3 senores or measuring blood oxygen level and heart rate; sensor module for ECG signal data; and an infrared sensor for body temperature.	Diagnosis physiological data, including blood oxygen levels, heart rates, body temperatures, and ECG signals, based on remote	CNN with attention layers	ECG Heartbeat Categorization Dataset - MITBIH Arrhythmia Database	98.2%
[28]	Remote model: Smartphone and wearable device.	Remote detection cardiac disease prediction.	XGBoost Bi-LSTM Model	UCI Heart Disease Data	99.4%
Proposed solution	Real-time model: Arduino board with 4 sensors: Heart rate and oxygen sensor human body temperature and vibration.	Early detection of health problems in a timely system.	LR	Real time dataset	91.97%
			SVM		96.38%
			PCA-SVM		97.87%
			RF		98.6%
			Hybrid system		99.86%

This tendency of performance reveals the fact that on the one hand, when applied individually, the models perform well showing their predictive potential, on the other hand, combining them in a logical hybrid system allows realizing their synergetic effect and making up the deficiencies of each model separately. Compared to the obvious accuracy enhancement of the hybrid model, the idea of ensemble-based decision mechanisms remains the most beneficial in complex classification problems involving health status prediction.

In order to assess the proposed approach against the prior approaches, the assessment is based on the methods, the dataset and the ML method. The traditional approach provides the best way of working, so as to get desired results, unlike the hybrid approach which achieves impeccable outcome as demonstrated in Table 6.

6. Conclusion

The paper proposes an IoT based remote health monitoring system for remote monitoring designed for early detection of health problems in home and clinical settings. The system architecture integrates biomedical sensors with an Arduino microcontroller to acquire physiological data, including blood oxygen saturation, heart rate, body temperature, and vibration signals.

The collected data were saved in virtual cloud environments as a Thing Speak server.

A key contribution of this work lies in the comparative evaluation of multiple machine learning models used to classify the general health status from the sensor readings such as heart rate, oxygen level, temperature, and vibration.

In addition to establishing a balanced dataset of normal and abnormal patterns, the performance of LR, SVM, PCA-SVM, and RF were compared. The findings showed that the proposed method was superior to conventional ones. For instance, the accuracy of the best-performing algorithm, RF, reported 98.6 % accuracy and pointed out its ability to handle only structured features with clear decision margins. The study has shown great potential for remote health monitoring by integrating deep learning techniques and the Internet of Things. By making the most of IoT sensors, real-time data can be obtained, while machine learning algorithms enable early detection of health problems.

In contrast to all prior work in real-time healthcare monitoring systems, the proposed model is built on real-time data, yet all machine learning models rely on public datasets for training and testing.

Future research should therefore focus on incorporating deep learning and temporal modelling techniques, such as recurrent or attention-based architectures, to better capture dynamic physiological patterns. Training and validating models using large-scale real-world datasets collected over extended periods would further enhance system reliability and generalisability. Moreover, integrating edge intelligence to enable on-device processing and reduce latency, as well as addressing security and privacy concerns in cloud-based health data transmission, represent important directions for future work. These enhancements would strengthen the applicability of IoT-enabled intelligent healthcare systems and contribute to their adoption in real-world clinical practice.

References:

[1]. Patel, B., & Shah, P. (2022). Operating system support, protocol stack with key concerns and testbed facilities for IoT: A case study perspective. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 5420-5434.

[2]. Fadhil, H. M., Hassan, H. S., & Abed, E. M. (2022). IoT E-health System for Cardiac Telemonitoring. *2022 International Conference on Computer and Applications (ICCA)*, 1-6.

[3]. Alizadeh, T. (2017). An investigation of IBM's Smarter Cites Challenge: What do participating cities want?. *Cities*, *63*, 70-80.

[4]. Chong, J. W., et al. (2018). Motion and noise artifact-resilient atrial fibrillation detection using a smartphone. *IEEE journal on emerging and selected topics in circuits and systems*, *8*(2), 230-239.

[5]. Ramya, R., & Ramamoorthy, S. (2024). Hybrid Fog-Edge-IoT architecture for real-time data monitoring. *International Journal of Intelligent Engineering and Systems*, *17*(1), 228-239.

[6]. Majeed, U., et al. (2021). Blockchain for IoT-based smart cities: Recent advances, requirements, and future challenges. *Journal of Network and Computer Applications*, *181*, 103007.

[7]. Mora, L., Deakin, M., & Reid, A. (2019). Strategic principles for smart city development: A multiple case study analysis of European best practices. *Technological Forecasting and Social Change*, *142*, 70-97.

[8]. Javed, A. R., et al. (2023). A survey of explainable artificial intelligence for smart cities. *Electronics*, *12*(4), 1020.

[9]. Biswas, S., et al. (2021). Cryptocurrency price prediction using neural networks and deep learning. *2021 7th international conference on advanced computing and communication systems (ICACCS)*, *1*, 408-413.

[10]. Thottempudi, P., et al. (2025). Digital health resilience: IoT solutions in pandemic response and future healthcare scenarios. *Discover Sustainability*, *6*(1), 144.

[11]. Chatterjee, A., et al. (2023). A systematic review and knowledge mapping on ICT-based remote and automatic COVID-19 patient monitoring and care. *BMC health services research*, *23*(1), 1047.

[12]. Javaid, M., & Khan, I. H. (2021). Internet of Things (IoT) enabled healthcare helps to take the challenges of COVID-19 Pandemic. *Journal of oral biology and craniofacial research*, *11*(2), 209-214.

[13]. Dziak, D., Jachimczyk, B., & Kulesza, W. J. (2017). IoT-based information system for healthcare application: design methodology approach. *Applied sciences*, *7*(6), 596.

[14]. Rahman, G., & Wen, C. C. (2022). Omega network pseudorandom key generation based on dna cryptography. *Applied Sciences*, *12*(16), 8141. Doi: 10.3390/app12168141.

[15]. Chaki, J., et al. (2022). Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review. *Journal of King Saud University-Computer and Information Sciences*, *34*(6), 3204-3225.

[16]. Rothman, M. J., Rothman, S. I., & Beals IV, J. (2013). Development and validation of a continuous measure of patient condition using the electronic medical record. *Journal of biomedical informatics*, *46*(5), 837-848.

[17]. Uddin, M., & Syed-Abdul, S. (2020). Data analytics and applications of the wearable sensors in healthcare: an overview. *Sensors*, *20*(5), 1379.

[18]. Nasr, M., et al. (2021). Smart healthcare in the age of AI: recent advances, challenges, and future prospects. *IEEE Access*, *9*, 145248-145270.

[19]. Singh, L. (2023). BFCNN: Multi-Disease Classification Bee Inspired CNN Model. *International Journal of Intelligent Engineering & Systems*, *16*(4).

[20]. Chaki, J., & Woźniak, M. (2023). A deep learning based four-fold approach to classify brain MRI: BTSCNet. *Biomedical Signal Processing and Control*, *85*, 104902.

[21]. Pandey, B., et al. (2022). A comprehensive survey of deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 5083-5099.

[22]. Liu, H., & Tripathy, R. K. (2025). Machine Learning and Deep Learning for Healthcare Data Processing and Analyzing: Towards Data-Driven Decision-Making and Precise Medicine. *Diagnostics*, *15*(8), 1051.

[23]. Krones, F., et al. (2025). Review of multimodal machine learning approaches in healthcare. *Information Fusion*, *114*, 102690.

[24]. Shaik, T., et al. (2023). Remote patient monitoring using artificial intelligence: Current state, applications, and challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(2), e1485.

[25]. Camps-Valls, G., Bioucas-Dias, J., & Crawford, M. (2016). A special issue on advances in machine learning for remote sensing and geosciences [from the guest editors]. *IEEE Geoscience and Remote Sensing Magazine*, *4*(2), 5-7.

[26]. Constant, N., et al. (2021). Data analytics for wearable IoT-based telemedicine. *Wearable sensors*, 357-378. Academic Press.

[27]. Khan, A., & Zubair, S. (2022). An improved multi-modal based machine learning approach for the prognosis of Alzheimer's disease. *Journal of King Saud University-Computer and Information Sciences*, *34*(6), 2688-2706.

[28]. Chakraborty, S. (2022). Multi-Disease Detection using Hybrid Machine Learning. *Sch J Eng Tech*, *10*, 271-278.

[29]. Islam, M. R., et al. (2023). Deep learning-based IoT system for remote monitoring and early detection of health issues in real-time. *Sensors*, *23*(11), 5204.

[30]. Alzakari, S. A., et al. (2024). Enhanced heart disease prediction in remote healthcare monitoring using IoT-enabled cloud-based XGBoost and Bi-LSTM. *Alexandria Engineering Journal*, *105*, 280-291.

[31]. Henderi, H., Wahyuningsih, T., & Rahwanto, E. (2021). Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *International Journal of Informatics and Information Systems*, *4*(1), 13-20.

[32]. Relangi, N. D. S. S. K., Chaparala, A., & Sajja, R. (2023). Effective Groundwater Quality Classification Using Enhanced Whale Optimization Algorithm with Ensemble Classifier. *International Journal of Intelligent Engineering and Systems*, *16*(1), 214-223.