https://doi.org/10.31449/inf.v49i1.6875 Informatica **49** (2025) 93–104 **93**

Real-Time Smart Healthcare System Based on Edge-Internet of Things and Deep Neural Networks for Heart Disease Prediction

Messaoud Hameurlaine[1], Abdelouhab Moussaoui[2], Mustapha Bensalah[3]
E-mail: hameurelaine.messaoud@univ-tissemsilt.dz, abdelouahab.moussaoui@univ-setif.dz, bensmusta@gmail.com
[1]LEIMA Laboratory, Center Universitaire Ahmed Ben Yahia El Wancharissi de Tissemsilt, Tissemsilt, Algeria
[2]Computer Science Department , Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria
[3]Signal Processing Laboratory, Military Polytechnic School, Algiers, Algeria

Keywords: Smart healthcare system, edge computing, internet of things, deep learning, deep neural network

Received: August 7, 2024

With technological advancements, smart health monitoring systems have become increasingly vital and popular. The rise of smart homes, appliances, and medical systems, along with the pivotal role of the Internet of Things (IoT), is significantly enhancing healthcare services by improving data processing and predictive capabilities. IoT not only aids in predicting heart disease but also supports emergency responses. However, traditional data transfer methods are inefficient in terms of time and energy, resulting in high latency and consumption. Edge computing, alongside deep learning methods, provides effective solutions with superior performance. This paper introduces a Real-Time Smart Healthcare System utilizing Edge-Internet of Things and Deep Learning. The primary objective of this system is to monitor patient health changes, predict heart disease, and automate medication administration in real time. The study presents a DNN-based prediction model that leverages edge computing and IoT. This model processes health data from IoT devices, while edge devices deliver timely health predictions to doctors and patients via edge and cloud servers. The proposed system is evaluated on performance parameters, demonstrating superior results compared to other methods. By integrating edge computing, IoT, and deep learning, this system enables efficient real-time health monitoring and prediction, benefiting both healthcare professionals and patients. It demonstrates exceptional performance with an accuracy of 96.15%, precision of 92.86%, recall (sensitivity) of 97.50%, and an F1-score of 94.87%.

Povzetek: Razvili smo sistem za pametno zdravstveno oskrbo, ki temelji na tehnologijah Edge-IoT in globokem učenju in napoveduje srčne bolezni v realnem času. Sistem uporablja globoko nevronsko mrežo za obdelavo podatkov iz IoT naprav, kar omogoča pravočasne zdravstvene napovedi.

1 Introduction

As the prevalence of heart disease patients continues to grow, the strain on current healthcare systems is on the rise. Addressing this challenge requires the increased involvement of specialists. However, it is crucial to act swiftly when dealing with heart patients, especially in emergencies. An efficient Smart Healthcare Surveillance System (SHSS) can effectively tackle these issues by offering a range of services, including monitoring, remote treatments, autonomous actions, and real-time situation management[1], [2]. Such a system can provide timely and effective support to heart disease patients, ensuring rapid and efficient responses in critical situations. Therefore, the demand for modern and future healthcare services requires computational power and rapid response times. In recent years, advances in information technology and artificial intelligence (AI) have played a transformative role across society, highlighting exponential progress in computational power, data storage, and electronic transmission capabilities [3], which are now critical to modern healthcare systems. The proliferation of AI-driven systems has unlocked the potential for unprecedented levels of automation and real-time data processing, especially through edge computing.

Nonetheless, mobile cloud computing (MCC), which preceded edge computing, encountered similar obstacles, such as high costs of data transmission, delayed response times and restricted network reach. Several investigations have contrasted cloud-centric and edge-centric computing, concluding that solely edge computing aligns with the demands for reduced latency, mobility, and energy conservation [4], [5]. In the healthcare sector, edge computing outperforms traditional cloud computing. Healthcare professionals can deliver remote medical assistance to individuals with chronic conditions through the utilization of wearable devices and ambient sensors to monitor vital signs. This is achievable thanks to the oversight and flexibility afforded by edge computing systems. Furthermore, doctors can detect patient risks based on sensor data, regardless of their location. For superior care delivery, it's essential that edge devices and nodes promptly perform data operations [6]. Real-time healthcare applications are also crucial for immediate risk detection and intervention. A Real-time Semantic Healthcare System is developed in[7] to identify Visual Risks for Elders and Children in surveillance videos, enabling swift responses to prevent accidents.

We propose a healthcare monitoring system that relies on Edge Internet of Things and deep learning technologies. The primary objectives of this system are as follows:

- An envisioned design for a Heart Disease Smart Healthcare Surveillance System integrates IoT-Edge-Cloud technology. This system harnesses the power of Edge devices to efficiently handle real-time situations and optimize performance by minimizing computation and transmission overhead.
- Data gathered from IoT sensors for predicting the risk of heart disease undergoes a sequence of preprocessing procedures and analysis at the Edge layer, facilitated by FPGA programmable processors. This process ensures that the data is prepared and analyzed efficiently and effectively within the system edge layer.

The subsequent segments of the paper are presented as follows: a comprehensive overview of existing research and studies related to the topic is presented in section 2. Section 3 offer detailed elaboration on our proposed architecture. The analysis of the results is presented in Section 4. Finally, Section 5 summarizes the findings and conclusions of the research.

2 Related work

In this section, we present a compilation of related works. One such research discusses an IoT-based healthcare system capable of monitoring and tracking patients, staff, and biomedical devices, while also handling emergency situations effectively[8]. Additionally, a framework is introduced that includes a real-time alert generation system to guarantee swift responses. Furthermore, there is the introduction of an IoT-cloud-based framework for healthcare applications, with a specific focus on real-time prediction of health vulnerabilities during workout sessions[9]. A healthcare system based on IoT is created to meet the demand for intelligent health monitoring. The framework introduces a combination of Fully Homomorphic Encryption (FHE) and machine learning [10]. This framework enables encrypted analysis of biosignal data, including aggregation, real-time monitoring, and abnormality detection.

In the healthcare industry, predictive analytics covers a wide range of techniques, ranging from conventional linear models to sophisticated machine learning techniques [11]. Among these techniques, deep learning (DL), a subset of ML, has proven to be highly reliable and robust. DL excels in automatically handling and learning from vast and complex healthcare datasets, providing valuable perspectives and efficient solutions to complex issues. Its application in diverse medical domains has consistently shown superior performance compared to classic designs. More precisely, the recurrent neural network (RNN) has demonstrated its effectiveness in managing prolonged dependencies in input data. RNN has become prominent in analyzing temporal events, particularly in applications that involve time-sequential data [12].

Several studies have focused on the diagnosis and predictive modeling of heart disease. Botros et al.[13] introduced two models for detecting heart failure from electrocardiogram signals: a Convolutional Neural Network and an enhanced version that includes an SVM layer, achieving over 99% accuracy, sensitivity, and specificity. This framework aids professionals and allows real-time processing with mobile equipments. The authors in [14] examined heart disease prediction using six ML models, as logistic regression and random forest. Logistic regression reached 90.16% accuracy on the Cleveland data, and AdaBoost achieved 90% on the IEEE Dataport data. The accuracy of the soft voting group classifiers was improved to 93.44% and 95%, respectively. Nancy et al.[2] utilized bidirectional LSTM for heart disease prediction, achieving an accuracy of 0.98 and outperforming existing methods. This highlights the importance of timely disease prediction for early intervention. Authors in [15] created a Smart Cardiovascular Disease Diagnostic Framework using Internet of Things devices. Their ConvNet and ConvNet-LSTM design successfully obtained a 98% accuracy rate in identifying atrial fibrillation using cloud architecture and DL. The combination of IoT devices and cloud computing with deep learning models offers transformative possibilities for healthcare, particularly in remote health monitoring and precise disease diagnosis. Table 1 present a summary of some recent approaches along with their respective performance and advantages.

Although the existing literature demonstrates promising approaches in heart disease prediction and healthcare monitoring, several limitations remain. Many of the techniques rely on powerful computational resources, such as deep learning models (CNN, LSTM), which can be costly and require large datasets for training. Additionally, real-time prediction remains a challenge, particularly in resource-constrained environments such as wearable IoT devices and edge computing systems. While IoT-based healthcare systems are increasingly utilized, their effectiveness often relies on centralized cloud processing, which introduces latency and scalability issues.

The proposed system addresses these challenges by leveraging edge computing and IoT integration, enabling real-time prediction and monitoring at the point of care, with significantly reduced latency. Our approach also integrates deep learning models such as DNN for improved accuracy, while operating efficiently on edge devices with limited resources, making it a valuable advancement over existing state-of-the-art (SOTA) methods in healthcare prediction systems.

Table 1: Summary of various approaches along with their respective performance and advantages.

Approach	Description	Advantages	Disadvantages	Reference
Meta classification technique	Uses multiple classifiers to improve prediction by combining their outputs	Combines strengths of different models for improved accuracy	Can be computationally expensive due to ensemble complexity	Latha et Jeeva (2019) [16]
Hybrid random forest with linear model	Combines random forest and linear models to predict heart disease	Balances complexity and interpretability	May not fully capture nonlinear relationships in data	Mohan et al. (2019) [17]
Statistical model and deep neural network	Combines traditional statistical methods with deep neural networks for heart disease prediction	Leverages both classic and modern techniques for robust results	Requires large datasets for deep learning to be effective	Moreno-Ibarra et al. (2019) [18]
Bi-directional LSTM (C-BiLSTM)	Uses BiLSTM to handle sequential data for heart disease prediction	Captures temporal dependencies in patient data, improving prediction accuracy	Computationally intensive, requires significant training data	Dileep et al. (2023) [19]
Hyperparameter tuning and cross-validation with ML	Employs hyperparameter optimization to enhance performance of machine learning models	Improves generalization ability of models	Can lead to overfitting if not carefully tuned	Ahmed et al. (2020) [20]
Random Forest	Utilizes Random Forest for heart disease classification	Easy to interpret, performs well with unstructured data	Can be less effective with highly imbalanced datasets	Dhanamjayulu et al. (2022) [21]
Optimized ensemble fuzzy ranking (OEFR)	Uses an ensemble of fuzzy ranking algorithms for heart disease prediction	Optimizes predictions, reduces error	May not handle very large datasets well	Managala et al. (2023) [22]
RNN (Recurrent Neural Network)	Uses RNNs for heart disease prediction from temporal data like ECG signals	Suitable for sequential data, handles time-series data well	Struggles with long-term dependencies in data	Almujally et al. (2023) [23]
IoT-Cloud-Based Healthcare System	Implements an IoT-based framework for real-time monitoring of heart disease during workouts	Integrates multiple data sources (physiological, behavioral) for holistic health tracking	Security concerns related to data transmission and privacy	Nancy et al.(2022) [2]
Deep Forest Cascade Technique	Uses a cascade of deep forest models to predict heart disease	High accuracy in prediction, interpretable output	Requires fine-tuning of cascade layers	Askar (2023) [24]
XGBoost	Uses an optimized gradient boosting model for classification tasks	Very high prediction accuracy, great for structured data	Sensitive to noisy data, requires proper feature engineering	Gracious et al. (2024) [25]
Optimized Random Forest with SMOTE	Combines Random Forest with SMOTE to balance class distributions for heart disease prediction	Improves model robustness and generalization	Computationally expensive and may require a lot of time to optimize	ishaq et al. (2023) [26]

3 Materials and methods

3.1 Architecture of the proposed smart healthcare system

IoT technology plays a pivotal role in various real-time applications, enabling seamless interaction between objects and individuals. However, the considerable volume of medical information produced by such equipment is a significant obstacle for the system, particularly in data storage, processing, and management. To address this challenge, we support the adoption of an intelligent system for heart disease diagnostic, employing edge IoT technology. The system, illustrated in Figure 1, aims to overcome the challenges posed by the massive data generated in healthcare settings. The architecture comprises 3 layers: the Cloud Layer, the Edge Layer, and the Data Generation Layer. The global architecture of the system is built upon the framework described in [27], but we have implemented several modifications. These include the integration of an intelligent sensor capable of automatically generating various physiological parameters continuously. The synergy between the National Instruments myRIO processor and a Wi-Fi module facilitates wireless data transmission to the cloud server [28]. Additionally, the system integrates real-time online monitoring of health status [6]. Moreover, the cloud component is employed for dispatching alert messages to patients, as documented in[21].

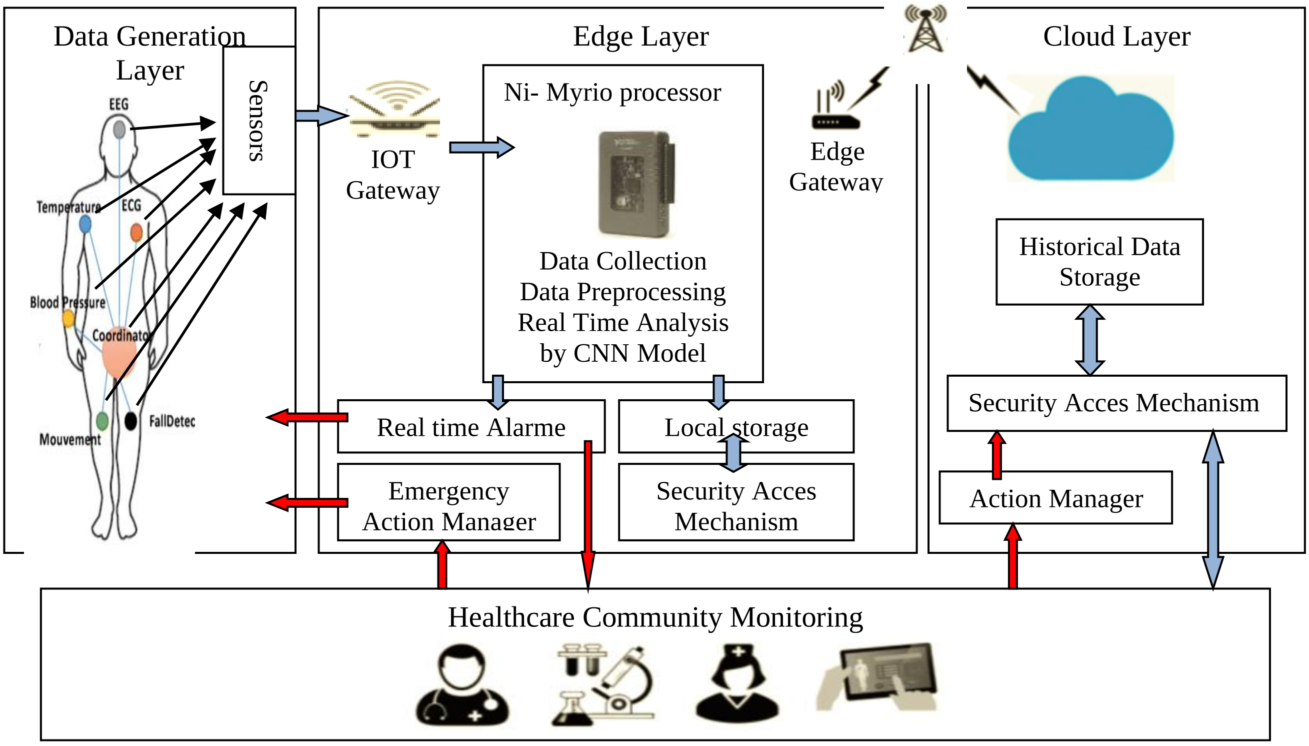

Figure 1: Proposed architecture of edge based system.

The proposed architecture enhances existing designs by incorporating specific adaptations tailored for real-time, resource-efficient processing on edge devices. Unlike traditional systems that rely solely on cloud computing for high-level data processing, our design emphasizes advanced local processing through edge computing, significantly reducing latency. The integration of intelligent sensors at the data generation layer provides continuous physiological parameter tracking, which is processed by a custom DNN model at the edge. This DNN model has been optimized for the constrained computational environment of edge devices, enabling it to operate effectively with lower energy consumption and processing power without compromising diagnostic accuracy.

3.2 Data generation layer

The function of this layer is to acquire physiological health parameters using a variety of Internet-connected or wearable devices. Typically, these devices have limited storage and computing capabilities. Local processing techniques can be implemented to address issues such as data redundancy, power consumption, and network transmission load. For connectivity, the data collection layer devices can establish a connection with smart phones or other mobile smart terminals via Bluetooth technology. To facilitate high-speed transmission of measured signals, all smart sensors are linked to an NI myRIO platform through a Wi-Fi module capable in swift data transfer. The NI myRIO platform utilizes a low-power Xilinx FPGA programmable processor, making it highly suitable for efficient signal transmission and reception tasks. To interact with the collected data, the NI-myRIO module communicates with an application that provides the capability to use a web-based interface for data visualization and analysis.

3.3 Edge layer

Primary role of this layer is to execute computations for early detection and take necessary actions based on the acquired physiological data. The Edge layer primarily consists of smart phones and other intelligent mobile terminals. It serves as network layer devices, enabling data communication functions of the medical IoT gateway. We use here NI myRIO. Moreover, they host application layer software, including preprocessing algorithms for filtering and consolidating data, thereby enhancing real-time analysis speed. The process of analysis is streamlined by deep learning models, offering reliability and precision in the detection of potential health concerns. Once the analysis is complete, the decision-making module, with the assistance of healthcare specialists, determines whether immediate action is required. If necessary, an alarm is raised to alert the healthcare community and an autonomous system capable of addressing emergencies in real-time. In instances where no alarm is triggered, both the data and analysis results are archived in the Edge layer before being transmitted to the Cloud layer. The access procedure incorporates appropriate access control measures for the healthcare community, granting authorized individuals the ability to retrieve and interact with the data as required.

3.4 Cloud layer

This layer focuses on data storage, latency-tolerant analysis, and access control for the healthcare community. It acts as a repository for collected data and allows for further analysis that can tolerate certain delays. The access mechanism ensures proper control over data access by authorized individuals. The analysis results are shared with the healthcare community, enabling timely actions based on the derived insights. The healthcare community can utilize cloud-based solutions for monitoring patients as well.

3.5 Security and confidentiality

Our smart healthcare system, which is built on the Internet of Things (IoT), incorporates a security solution provided by [23]. Our system complies with security standards like HIPAA and GDPR by using encryption and secure authentication methods, ensuring that only authorized users and devices can access sensitive health data while protecting user privacy. This solution utilizes Zigbee and Firebase IoT authentication. During the transmission of health information, a 128-bit encryption is utilized to secure the JSON file as a token. Firebase cloud functions authenticate the officer's device token by generating a custom token with precise credentials and claims. Both the 128-bit device token and the Firebase custom token serve as authentication mechanisms for real-time data exchanges. After the user's identity is verified, the authorization process makes use of Firebase's universal security standards. The security system consists of three steps:

- The equipment identifier authenticates the request as being from permitted equipment, but it does not provide any important details for identifying the owner.
- A customized certificate includes user identification but does not have profile details and cannot be automatically recognized by Firebase servers due to potential revocation or key rotation.
- The signInWithCustomToken API verifies the claims of the custom token, and then the backend produces a Firebase Identity token. This token, which includes the user's characteristics, serves as indisputable evidence of authorization and remains valid for duration of one hour.

3.6 Proposed deep learning model

In the past few years, there has been significant research and extensive implementation of deep learning algorithms, aimed at extracting valuable information from different varieties of data. Various deep learning architectures have been implemented to accommodate the diverse characteristics of input data, encompassing conventional neural networks, deep neural networks, and recurrent neural networks.

In this particular case, a deep neural network model has been adopted and modified to predict heart disease. The specific architecture of the DNN is illustrated in Figure 2. The described sequential model architecture is a dense neural network (DNN) comprising multiple fully connected layers with regularization mechanisms to enhance generalization and prevent over fitting. The model begins with a dense layer of 128 neurons, using the ReLU (Rectified Linear Unit) activation function. The weights in this layer are initialized with a normal distribution, and L2 regularization with a coefficient of 0.001 is applied to mitigate over fitting by penalizing large weights. This layer is followed by a Dropout layer that randomly drops 20% of the neurons during training, providing further regularization. This structure of dense and dropout layers is repeated with 64, 32, and 16 neurons in subsequent dense layers, each maintaining the same initialization and regularization techniques. The model continues with another Dropout layer after each dense layer to ensure regularization is consistently applied throughout the network. Following these, a dense layer with 8 neurons using the softmax activation function is included. The softmax activation function is typically used for multi-class classification and outputs a probability distribution classes. Another Dropout layer follows this, and finally, the model includes a dense layer with 2 neurons and a sigmoid activation function. The sigmoid activation function is often used for binary classification tasks, outputting probabilities for two classes.

The model is trained with the Adam optimizer, employing a learning rate of 0.001. The chosen loss function is categorical crossentropy, so it is well-suited for issues involving multi-class categorization. The categorical crossentropy loss function measures the dissimilarity between the true labels and the predicted probabilities, penalizing incorrect classifications more severely. The formula for categorical crossentropy loss is:

$$Loss = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \qquad (1)$$

where N is the number of classes, y_i is the true label for class i , and $\hat{y}_i$ is the predicted probability for class i.

L2 regularization, often referred to as Ridge regularization, is a method employed in machine learning models to mitigate over fitting. It achieves this by incorporating a penalty term into the loss function. The "L2" refers to the L2 norm, which is the sum of the squared values of the weights. This penalty term discourages the model from fitting the noise in the training data by shrinking the weight values, thus promoting simpler models with smaller weights.

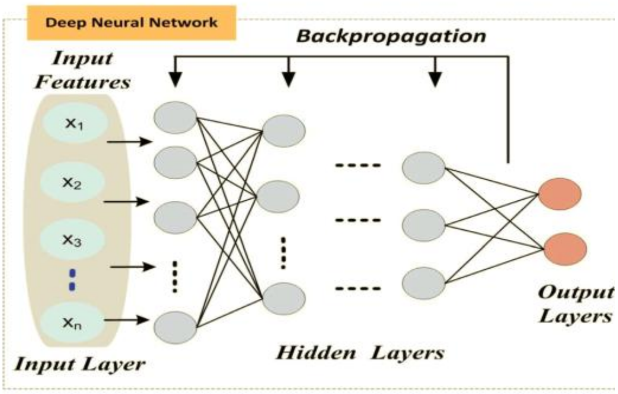

Figure 2: DNN model

Mathematically, L2 regularization adds the following term to the loss function:

$$P = \lambda \sum_i \omega_i^2 \quad (2)$$

Where λ is a regularization parameter that controls the strength of the penalty, and ω_i are the weights of the model.

When the loss function with L2 regularization is minimized, the model not only tries to minimize the original loss but also tries to keep the weights small. This helps in preventing over fitting, as models with smaller weights are less likely to fit the noise in the training data and more likely to generalize better to new, unseen data.

The training phase entails the adjustment of the model to the test data with a validation split using the same test data over 60 epochs and a batch size of 15. The model's training was optimized by employing an early stopping criterion, where training is terminated if accuracy stabilizes without improvement over a set number of epochs, ensuring optimal model performance while avoiding unnecessary computation. Other hyperparameters were determined experimentally to achieve the most effective configuration. For instance, the batch size was set to 15 to maintain a balance between memory efficiency and gradient updates, which aids in faster convergence. The learning rate for the Adam optimizer was chosen as 0.001 after testing different rates and selecting the one that yielded the best training and validation accuracy. Moreover, each layer's dropout rate and the L2 regularization coefficient were tuned to prevent overfitting effectively. These hyperparameter values were fine-tuned by monitoring model performance across multiple experimental runs, adjusting values iteratively to achieve a reliable, accurate prediction model for heart disease diagnosis. Upon training completion, the model's performance is evaluated on the test data. The evaluation provides a loss and accuracy metric. Construction and training of the deep learning model took place on Edge layer or cloud server. The output layer of this model generates health assessment results, which are categorized into binary classes: 0 for health and 1 for illness.

3.7 Dataset

The Erbil Heart Disease Dataset [29], sourced from the Medical Help Centre, a specialized heart hospital in Erbil, Iraq, contains data on 333 patients, each with 21 attributes as detailed in Table 2. This publicly available dataset aims to facilitate the prediction of heart disease using information specific to the local population. The dataset's attributes are organized into five categories: demographic information, medical history, physical examinations and symptoms, medical laboratory tests, and diagnostic features. These attributes were carefully chosen based on expert medical advice to ensure their relevance and significance for heart disease prediction. By leveraging this dataset, researchers can develop and refine predictive models that are tailored to the unique characteristics of the patients from this region, ultimately enhancing the accuracy and effectiveness of heart disease diagnosis and treatment strategies. Figure 3 illustrates the distribution of the target variable for heart disease. It visually represents the proportion of patients with and without heart disease, aiding in understanding the dataset's balance and the prevalence of the condition.

Table 2: Data description

Attribute	Description
Age	Patients' ages, measured in years.
Sex	The patient's gender is indicated by a value of 1 for female and 0 for male.
Smoking	Indicate whether the patient is a smoker or not (0=No, 1=Yes)
Years	Duration of smoking for smokers
LDL	The patient's LDL-Cholesterol ratio
Chp	Types of chest pain are categorized as follows: 1= Typical angina, 2= Atypical angina, 3= Non-anginal pain, and 4= Asymptomatic.
Height	The patient's height, measured in centimeters.
Weight	The patients' weight, measured in kilograms.
FH	History of heart disease among family members
Active	Indicate the patient's level of activity (0=Inactive, 1=Active)
Lifestyle	Residence location: 1 = City, 2 = Town, 3 = Village
CI	Has the patient undergone any cardiac catheterization or any invasive procedures involving the heart? (0 indicates No, while 1 indicates Yes)
HR	Cardiac pulse ratio
DM	Presence of diabetes: 0 = No, 1 = Yes
Bpsys	Ratio of systolic blood pressure
Bpdias	Ratio of diastolic blood pressure
HTN	The patient's hypertension status: 0 for "No," 1 for "Yes."
IVSD	Echo parameter: Interventricular Septal Thickness during Diastole (IVSD), a measurement utilized in determining Left Ventricular Hypertrophy (LVH).
ECGpatt	The ECG test includes four categories: ST-Elevation (1), ST-Depression (2), T-Inversion (3), and Normal (4).
Qwave	Presence of the Q wave: 0 for "No," 1 for "Yes."
Target	The patient's heart disease status: 0 for "without heart disease," 1 for "with heart disease."

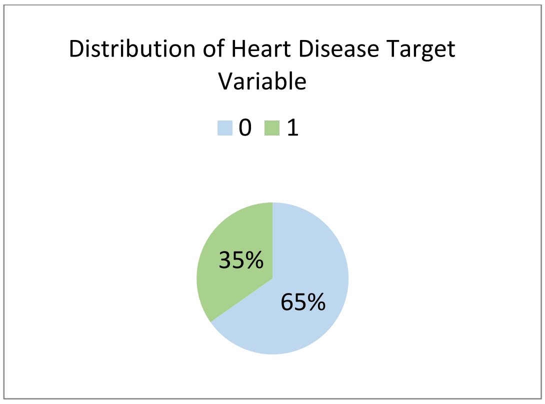

Figure 3: Distribution of target attribute

3.8 Data preprocessing

Data preprocessing plays a pivotal role in building a more precise machine learning model. This phase encompasses several tasks geared towards enhancing data quality, including the identification and management of missing values, the detection and elimination of outliers, and the selection of pertinent features. In addition to the preprocessing operations already performed on the Erbil data, we have implemented the following supplementary preprocessing steps to enhance their usability further.

3.8.1 Normalization

Normalization, particularly standardization, is a crucial preprocessing step in data analysis and machine learning. It adjusts the scale of data to ensure that each feature contributes equally, preventing those with larger scales from dominating results. This process involves centering the data around zero by subtracting the mean of each feature, effectively neutralizing biases introduced by different scales. Subsequently, the data is scaled by dividing it by the standard deviation of each feature, normalizing variance across features. This transformation, which results in each feature having a mean of zero and a standard deviation of one, is beneficial for algorithms sensitive to feature scales, such as linear regression and neural networks. Standardization accelerates gradient descent convergence and enhances model performance, ensuring equitable feature treatment and improving analysis or model training accuracy.

3.8.2 One-hot encoding

The operation of converting class labels into a categorical format, known as one-hot encoding, transforms numerical class labels into binary vectors. This process is essential in deep learning as it ensures that each class label is represented equally and without any implicit ordinal relationship. One-hot encoding helps neural networks to interpret the labels correctly, facilitating accurate computation of loss functions like categorical cross-entropy. This encoding method boosts the model's capacity to discern patterns from the data efficiently, thereby enhancing its performance in classification tasks.

3.8.3 Identification of missing values

This operation is performed to verify the absence of missing values in a dataset. By calculating the total percentage of missing data, we can confirm whether the dataset is complete. The process involves counting all missing entries across the dataset, ensuring that the dataset is fully intact and reliable for subsequent analysis or model training without requiring further imputation or cleaning steps.

4 Results and Discussion

4.1 Training and testing performance

The model was trained using 80% of the available data, with the remaining 20% set aside for testing. The assessment of the models includes the utilization of different performance measures, encompassing accuracy, specificity, F1 score, precision, and the recall. Accuracy is a measure used to assess the predictive ability of a DL model through the comparison of the expected outcome with the actual outcome. In the context of predicting heart disease, the classifier's ability to precisely determine the existence or non-existence of the disease is assessed through the true positive (TP) and true negative (TN) values. Conversely, false positive (FP) and false negative (FN) values highlight the inaccuracies in the models' predictions. Precision gauges the ratio of observed positive instances among all the predicted positive instances. Recall, alternatively known as sensitivity or the true positive rate, computes the ratio of actual positive instances correctly identified by the model. Specificity, conversely, evaluates the ratio of all negative instances that the model accurately predicts. Meanwhile, the F1 score is a metric that amalgamates precision and recall, delivering a harmonized measure of the model's performance. It computes the harmonic mean of precision and recall, assigning equal significance to both metrics.

Figure 4 illustrates the results obtained during the training and testing phase of the proposed model. This figure highlights a crucial role in evaluating and analyzing the performance of the proposed model, providing insights into its training and testing performance, respectively. It provides the value of accuracy and loss that evaluate the efficacy of the model and its performance on the unobserved test data.

By examining Figure 4, one can gain insights into how well the model performs on the test data, allowing for a comprehensive understanding of its performance and the potential impact on real-world scenarios. The proposed system demonstrates exceptional performance with an accuracy of 96.15%, precision of 92.86%, recall (sensitivity) of 97.50%, and an F1-score of 94.87%.The accuracy achieved for the proposed model was 96.15%. This indicates that the model's predictions aligned with the desired output in approximately 96.15% of cases. A high accuracy value like this suggests that the model performed exceptionally well in accurately classifying instances and making correct predictions.

4.2 Comparative analysis

The evaluation of the suggested work centers on assessing prediction accuracy through the application of state-of-the-art approaches to heart disease datasets. A comparative analysis has been conducted out to analyze the accuracy results of the proposed model in comparison to existing models documented in the literature (Table 1).

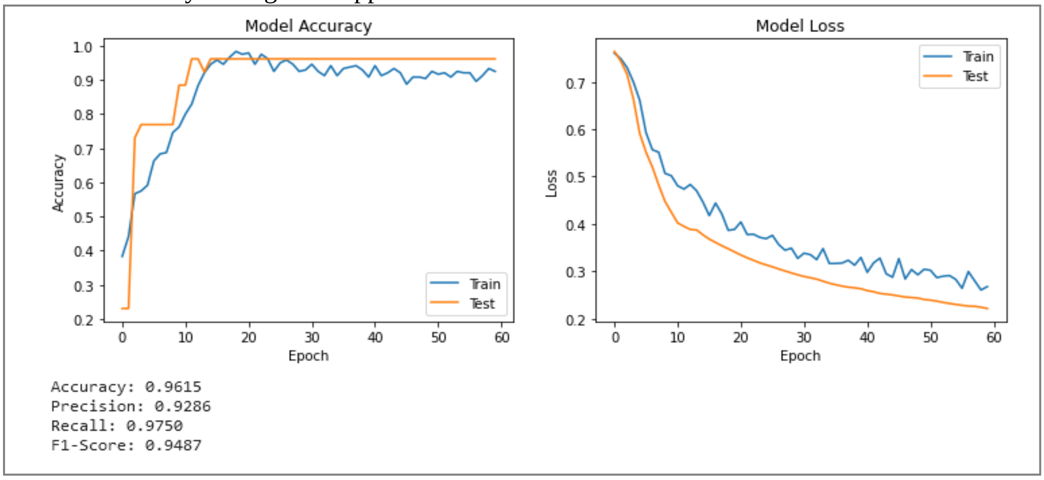

Figure 4: Training and testing performance: accuracy and loss across epochs

The outcomes of this comparison are illustrated in Table 3. In our research, we propose a deep learning model that achieves an accuracy of 96.15%, demonstrating its superior performance compared to several established methods. Specifically, our model outperforms the Meta classification technique, which achieves an accuracy of 85.48%, and the Hybrid random forest with a linear model, which reaches 88.70%. It also surpasses the Statistical model and deep neural network with 91.57% accuracy and the bi-directional LSTM (C-BiLSTM) algorithm, which attains 94.78%. Although hyperparameter tuning and cross-validation with machine learning yield 94.90% accuracy, and the traditional Random Forest achieves 95.25%, our model still exhibits better performance. Other techniques, such as CNN (93.98%), MLP (91.20%), RNN (91.00%), and the Extra Tree Classifier with SMOTE (92.62%), also fall short in comparison. The Deep Forest Cascade Technique (92.56%) and Random Forest without SMOTE (88.89%) further underscore the effectiveness of our proposed approach.

Despite the high accuracy achieved by the Optimised Ensemble Fuzzy Ranking (OEFR) strategy (96.72%), the LSTM model (96.91%), and the Fuzzy Information System with Bi-LSTM (98.86%), our approach presents several notable advantages. While these models demonstrate impressive results, they are often trained on datasets containing only 14 attributes, which limits the scope of patient information considered. In contrast, our model is trained on a dataset with 21 attributes, providing a more comprehensive analysis by integrating additional health parameters that are critical for precise heart disease prediction. Furthermore, the aforementioned methods do not account for real-time constraints, nor do they consider training time, which is essential for practical deployment in dynamic healthcare environments. Our approach emphasizes real-time prediction, leveraging edge computing to process data swiftly and effectively, making it better suited for real-world applications where timely responses are crucial. Thus, while some models may have higher reported accuracy, our model's comprehensive feature set and real-time capability underscore its suitability and effectiveness in practical healthcare settings.

The outcomes of the comparison with previously conducted related studies demonstrate that the proposed system outperforms these systems. With the increasing importance of real-time smart systems in healthcare, which heavily rely on IoT technology, tasks such as rapid processing become critical as they require minimal delays and are context-sensitive.

4.3 Training time comparison

To provide a more comprehensive time comparison, it's important to include the hardware specifications of the environment in which the experiments were conducted. In this study, the computations were performed using Google Colab, which utilizes a cloud-based GPU environment. Specifically, the model was run on a Tesla T4 GPU with a standard CPU configuration, offering a balance of performance and accessibility. The comparison of training times for various models, including transfer learning models, the CNN model proposed in [23], and the newly proposed DNN model, is shown in Table 4. This comparison highlights their computational complexity. Despite the variations in training conditions, such as the data used and the hardware, the newly proposed DNN model demonstrates a significantly shorter training time compared to the other models. This efficiency is particularly crucial for real-time processing, as it allows for faster model updates and deployment, which is essential in dynamic environments. The reduced training time, combined with the model's high accuracy, makes it especially suitable for edge computing applications. Implementing this model on edge devices enables real-time analysis and decision-making, thereby enhancing its practicality and effectiveness in real-world scenarios. The efficient processing and deployment capabilities of the proposed model highlight its advantages in applications requiring immediate results and timely responses.

Table 3: A comparative examination of the proposed model against existing models.

Approach	Acc	Author-Year-Reference
Meta classification technique	85.48	Latha et Jeeva (2019) [16]
Hybrid random forest with a linear model	88.70	Mohan et al. (2019) [17]
Statistical model and deep neural network	91.57	Moreno-Ibarra et al. (2019) [18]
bi-directional LSTM (C-BiLSTM) algorithm	94.78	Dileep et al. (2023) [19]
Hyperparameter tuning and cross-validation with machine learning	94.90	Ahmed et al. (2020) [20]
Random Forest	95.25	Dhanamjayulu et al. (2022) [21]
Optimised ensemble fuzzy ranking (OEFR) strategy	96.72	Managala et al. (2023) [22]
CNN	93.98	Almujally et al. (2023) [23]
MLP	91.20	Almujally et al. (2023) [23]
RNN	91.00	Almujally et al. (2023) [23]
LSTM	96.91	Almujally et al. (2023) [23]
Random Forest without SMOTE	88.89	ishaq et al. (2023) [26]
Extra Tree Classifier with SMOTE	92.62	ishaq et al. (2023) [26]
XGBoost	93.26	Gracious et al. (2024) [25]
Deep Forest Cascade Technique	92.56	Askar (2023) [24]
Fuzzy information system and Bi-LSTM	98.86	Nancy et al.(2022) [2]
Proposed approch	96.15	

Table 4: The comparison of training times

Model	Training time
AlexNet (transfer learning)	32 min
VGG-16 (transfer learning)	29 min
CNN	24 min
Proposed	12 sec

4.4 Advantages of edge architecture

We have proposed an edge architecture model combined with a deep learning framework for heart disease prediction, achieving remarkable performance. The exponential proliferation of devices and the resulting surge in data traffic have significantly increased bandwidth consumption and service disruptions. The traditional cloud model struggles with issues like latency, bandwidth utilization, and connectivity, making it insufficient to handle these challenges alone. Our decentralized edge computing model addresses these limitations by processing and storing data close to the source. This proximity allows for efficient handling of vast IoT data using AI tools at the edge layer, significantly reducing latency and managing the substantial data volume from IoT devices. By integrating a hierarchical edge-fog-cloud architecture, our model enhances performance and reliability in heart disease prediction, leveraging the advantages of edge and fog computing to deliver superior predictive capabilities in healthcare applications.

4.5 K-Fold Cross-validation

To validate our proposed model, we applied 10-fold cross-validation using a heart failure dataset. The results presented in Table 5 demonstrate the model's robust performance, with an average accuracy of 96.99%. The precision, recall (sensitivity), and F1-score achieved are 96.30%, 97.33%, and 96.63%, respectively.

Table 5: 10 cross-validation results

Number of Fold	Accuracy	Precision	Sensitivity (Recall)	F1-Score
Fold 1	96.30	95.45	97.06	96.10
Fold 2	100	100	100	100
Fold 3	92.59	90	94.74	91.67
Fold 4	96.30	95.45	97.06	96.10
Fold 5	100	100	100	100
Fold 6	96.30	97.37	94.44	95.71
Fold 7	96.15	95.00	97.06	95.85
Fold 8	100	100	100	100
Fold 9	96.15	96.88	95.45	96.01
Fold 10	96.15	92.86	97.50	94.87
Average	96.99	96.30	97.33	96.63

These metrics indicate that the model consistently performs well across different folds. Specifically, several folds, such as Fold 2, Fold 5, and Fold 8, achieved perfect scores (1.00) across all metrics, highlighting the model's capability to accurately classify patient data.

Other folds, such as Fold 1, Fold 4, and Fold 6, also showed high performance with accuracy and F1-scores above 96%.

In our k-fold cross-validation study, we enhance the statistical interpretation presented in Table 6 by calculating the standard deviation and 95% confidence intervals for each metric. The standard deviation provides insight into the variability or spread of the model's performance across the folds, while the 95% confidence intervals give a range within which the true metric value is likely to fall, offering a measure of reliability and stability.

Table 6: Statistical interpretation of k-fold cross-validation study

Metric	**Mean**	**Stand. Deviat.**	**95% Confidence Interval Lower Bound**	**95% Confidence Interval Upper Bound**
Accur.	96.99	2.24	95.61	98.38
Precis.	96.30	3.12	94.37	98.23
Sensit.	97.33	2.01	96.09	98.57
F1-scor.	96.63	2.53	95.06	98.20

The standard deviation values reflect the consistency of the model's performance. Smaller standard deviations, such as for accuracy (2.24%), indicate more stable results across folds, while larger values, like for precision (3.12%), suggest slightly higher variability in the model's precision across different data splits. The calculated 95% confidence intervals for each metric indicate the consistency and stability of the model's performance across different folds. For accuracy, the interval is between 95.61% and 98.38%, showing minimal variation. Precision's confidence interval spans from 94.37% to 98.23%, suggesting slightly more variability. Sensitivity has an interval from 96.09% to 98.57%, demonstrating reliable detection of true positives with little fluctuation. The F1-score ranges from 95.06% to 98.20%, confirming a robust balance between precision and recall. These intervals reflect stable and consistent model performance.

These results affirm the effectiveness of the proposed edge architecture combined with a deep learning model in predicting heart disease, significantly outperforming traditional methods by leveraging the advantages of edge computing to process data efficiently and accurately.

5 Conclusion

This work introduces a model for a Real-Time Smart Healthcare System, tailored for predicting the risk of heart disease, leveraging Edge-IoT and DL technologies. The architecture consists of three layers, each with its required components, and employs a deep learning model for the task of prediction. The suggested system demonstrates outstanding performance, boasting average accuracy, precision, sensitivity, and F1-score values of 96.99%, 96.30%, 97.33%, and 96.63, respectively, surpassing other current models for predicting heart disease. However, this constitutes just one facet of the continual healthcare research using predictive analytics, with deep learning models holding tremendous untapped potential.

The model can be improved to autonomously generate tailored diet and exercise suggestions, taking into account an individual's health condition and guidance from a heart specialist. In this envisioned intelligent system for predicting heart disease, IoT devices are utilized for data acquisition, while edge computing manages data analysis, with the cloud reserved for other essential tasks. Nevertheless, our study acknowledges certain limitations, including the necessity for real-world validation, addressing privacy concerns, and ensuring scalability and compatibility within diverse healthcare infrastructures. Continued exploration and refinement of real-time smart healthcare systems are imperative for realizing their full potential in transforming healthcare delivery and improving patient outcomes.

To support real-world deployment, future work will focus on several concrete steps. These include rigorous testing in clinical environments to ensure model robustness and compliance with healthcare standards such as HIPAA or GDPR for data privacy. Enhancements to the system could allow for personalized diet and exercise recommendations tailored to individual health profiles, under medical guidance. Additionally, we plan to incorporate real-time patient feedback and specialist input, enabling the model to learn from real-world cases and improve over time.

The healthcare sector's effectiveness can undergo a transformation through accurate and timely disease forecasts, facilitating real time responses and smart decision-making by healthcare professionals, especially when leveraging fog/edge computing technologies. This integration has the potential to enhance the overall quality-of-service and revolutionize the healthcare industry.

Declarations

Competing interests

The authors affirm that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Ethical and informed consent for data used

The Erbil Heart Disease dataset (https://www.kaggle.com/datasets/hangawqadir/erbil-heart-disease-dataset) adheres to strict ethical guidelines and informed consent protocols to ensure the appropriate use and handling of data.

References

[1] M. Hartmann, U. S. Hashmi, and A. Imran, "Edge computing in smart health care systems: Review, challenges, and research directions," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 3, p. e3710, 2022, https://doi.org/10.1002/ett.3710.

[2] A. A. Nancy, D. Ravindran, P. M. D. Raj Vincent, K. Srinivasan, and D. Gutierrez Reina, "IoT-Cloud-Based Smart Healthcare Monitoring System for Heart Disease Prediction via Deep Learning," *Electronics*, vol. 11, no. 15, Art. no. 15, Jan. 2022, https://doi.org/doi: 10.3390/electronics11152292.

[3] M. Gams and T. Kolenik, "Relations between electronics, artificial intelligence and information society through information society rules," *Electronics*, vol. 10, no. 4, p. 514, 2021, Accessed: Nov. 10, 2024. [Online]. Available: https://www.mdpi.com/2079-9292/10/4/514

[4] S. Vimal, Y. H. Robinson, S. Kadry, H. V. Long, and Y. Nam, "IoT based smart health monitoring with CNN using edge computing," *Journal of Internet Technology*, vol. 22, no. 1, pp. 173–185, 2021. [Online]. Available: https://jit.ndhu.edu.tw/article/view/2471.

[5] X. Wang and Y. Li, "Edge Detection and Simulation Analysis of Multimedia Images Based on Intelligent Monitoring Robot," *Informatica*, vol. 48, no. 5, Art. no. 5, Mar. 2024, https://doi.org/doi: 10.31449/inf.v48i5.5366.

[6] X. Wu, C. Liu, L. Wang, and M. Bilal, "Internet of things-enabled real-time health monitoring system using deep learning," *Neural Comput & Applic*, Sep. 2021, https://doi.org/doi: 10.1007/s00521-021-06440-6.

[7] M. Belkebir, T. M. Maarouk, and B. Nini, "Real-time Semantic Healthcare System: Visual Risks Identification for Elders and Children," *Informatica*, vol. 48, no. 14, Art. no. 14, Sep. 2024, https://doi.org/doi: 10.31449/inf.v48i14.6271.

[8] G. S. Aujla and A. Jindal, "A Decoupled Blockchain Approach for Edge-Envisioned IoT-Based Healthcare Monitoring," *IEEE J. Select. Areas Commun.*, vol. 39, no. 2, pp. 491–499, Feb. 2021, https://doi.org/doi: 10.1109/JSAC.2020.3020655.

[9] M. Bhatia and S. K. Sood, "A comprehensive health assessment framework to facilitate IoT-assisted smart workouts: A predictive healthcare perspective," *Computers in Industry*, vol. 92–93, pp. 50–66, Nov. 2017, https://doi.org/doi: 10.1016/j.compind.2017.06.009.

[10] A. Alabdulatif, I. Khalil, X. Yi, and M. Guizani, "Secure Edge of Things for Smart Healthcare Surveillance Framework," *IEEE Access*, vol. 7, pp. 31010–31021, 2019,https://doi.org/doi: 10.1109/ACCESS.2019.2899323.

[11] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Brief Bioinform*, vol. 19, no. 6, pp. 1236–1246, Nov. 2018, https://doi.org/doi: 10.1093/bib/bbx044.

[12] S. K. Pandey and R. R. Janghel, "Recent Deep Learning Techniques, Challenges and Its Applications for Medical Healthcare System: A Review," *Neural Process Lett*, vol. 50, no. 2, pp. 1907–1935, Oct. 2019, https://doi.org/doi: 10.1007/s11063-018-09976-2.

[13] J. Botros, F. Mourad-Chehade, and D. Laplanche, "CNN and SVM-Based Models for the Detection of Heart Failure Using Electrocardiogram Signals," *Sensors*, vol. 22, no. 23, Art. no. 23, Jan. 2022, https://doi.org/doi: 10.3390/s22239190.

[14] N. Chandrasekhar and S. Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization," *Processes*, vol. 11, p. 1210, Apr. 2023, https://doi.org/doi: 10.3390/pr11041210.

[15] S. Tiwari *et al.*, "A smart decision support system to diagnose arrhythymia using ensembled ConvNet and ConvNet-LSTM model," *Expert Systems with Applications*, vol. 213, p. 118933, Mar. 2023, https://doi.org/doi: 10.1016/j.eswa.2022.118933.

[16] C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," *Informatics in Medicine Unlocked*, vol. 16, p. 100203, Jan. 2019, https://doi.org/doi: 10.1016/j.imu.2019.100203.

[17] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019,https://doi.org/doi: 10.1109/ACCESS.2019.2923707.

[18] M.-A. Moreno-Ibarra, Y. Villuendas-Rey, M. D. Lytras, C. Yáñez-Márquez, and J.-C. Salgado-Ramírez, "Classification of diseases using machine learning algorithms: A comparative study," *Mathematics*, vol. 9, no. 15, p. 1817, 2021. [Online]. Available: https://www.mdpi.com/2227-7390/9/15/1817

[19] P. Dileep *et al.*, "An automatic heart disease prediction using cluster-based bi-directional LSTM (C-BiLSTM) algorithm," *Neural Comput & Applic*, vol. 35, no. 10, pp. 7253–7266, Apr. 2023, https://doi.org/doi: 10.1007/s00521-022-07064-0.

[20] H. Ahmed, E. M. G. Younis, A. Hendawi, and A. A. Ali, "Heart disease identification from patients' social posts, machine learning solution on Spark," *Future Generation Computer Systems*, vol. 111, pp. 714–722, Oct. 2020, https://doi.org/doi: 10.1016/j.future.2019.09.056.

[21] C. Dhanamjayulu, G. V. Suraj, M. Nikhil, R. Kaluri, and S. Koppu, "A Machine Learning Algorithm-Based IoT-Based Message Alert System for Predicting Coronary Heart Disease," in *Advancements in Smart Computing and Information Security*, S. Rajagopal, P. Faruki, and K. Popat, Eds., in Communications in Computer and Information Science. Cham: Springer Nature Switzerland, 2022, pp. 362–376. https://doi.org/doi: 10.1007/978-3-031-23092-9_29.

[22] N. V. L. M. K. Munagala, L. R. R. Langoju, A. D. Rani, and D. V. rama K. Reddy, “Optimised ensemble learning-based IoT-enabled heart disease monitoring system: an optimal fuzzy ranking concept,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 0, no. 0, pp. 1–17, Jan. 2023, https://doi.org/doi: 10.1080/21681163.2022.2162439.

[23] N. A. Almujally *et al.*, “Monitoring acute heart failure patients using internet-of-things-based smart monitoring system,” *Sensors*, vol. 23, no. 10, p. 4580, 2023, Accessed: Jun. 01, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/23/10/4580

[24] S. K. Askar, “Deep Forest Based Internet of Medical Things System for Diagnosis of Heart Disease,” *ARO-THE SCIENTIFIC JOURNAL OF KOYA UNIVERSITY*, vol. 11, no. 1, Art. no. 1, Mar. 2023, https://doi.org/doi: 10.14500/aro.11174.

[25] G. S and S. C P, “ML based Age Related Heart Disease Prediction,” in *2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC - ROBINS)*, Apr. 2024, pp. 312–316. https://doi.org/doi: 10.1109/ICC-ROBINS60238.2024.10533975.

[26] A. Ishaq *et al.*, “Improving the Prediction of Heart Failure Patients’ Survival Using SMOTE and Effective Data Mining Techniques,” *IEEE Access*, vol. 9, pp. 39707–39716,2021,https://doi.org/doi:10.1109/ACCESS.2021.3064084.

[27] S. Basak and K. Chatterjee, “Smart Healthcare Surveillance System Using IoT and Machine Learning Approaches for Heart Disease,” in *Advancements in Smart Computing and Information Security*, S. Rajagopal, P. Faruki, and K. Popat, Eds., in Communications in Computer and Information Science. Cham: Springer Nature Switzerland, 2022, pp. 304–313. https://doi.org/doi: 10.1007/978-3-031-23092-9_24.

[28] P. Rajan Jeyaraj and E. R. S. Nadar, “Smart-Monitor: Patient Monitoring System for IoT-Based Healthcare System Using Deep Learning,” *IETE Journal of Research*, vol. 68, no. 2, pp. 1435–1442, Mar. 2022, https://doi.org/doi:10.1080/03772063.2019.1649215.

[29] “Erbil Heart Disease Dataset.” Accessed: Jun. 05, 2024. [Online]. Available: https://www.kaggle.com/datasets/hangawqadir/erbil-heart-disease-dataset