Transformer-Based Framework for Predictive Patient Monitoring using Multi-Modal Data

Saef Thallal[1]*

[1]Department of Computers Techniques Engineering, College of Technical Engineering, The Islamic University, Najaf, Iraq

Abstract. Recently, predictive patient monitoring has played a crucial role in Intensive Care Units (ICUs), and the early detection of clinical weakening saves lives and removes hospital readmissions. However, the existing Artificial Neural Network with Particle Swarm Optimization (ANN-PSO) approach fails to capture complex temporal patterns in multi-modal data and struggles to integrate static patient details. This limits their ability to produce accurate and timely risk predictions of critical outcomes. To overcome these challenges, a transformer-based predictive framework for patient monitoring was proposed. Initially, data were collected from the MIMIC-III dataset and processed using cleaning to remove inconsistencies, normalization to scale numerical features, and segmentation to fixed-length time windows. These segmented data are then fed to the transformer model, which is designed to capture the complex temporal patterns. Here, each time step in a patient's segment is transformed to embedding, and positional encoding is added to understand the order of events. These sequences are processed using a multihead self-attention mechanism, which enables the transformer to capture important temporal dependencies and feed-forward to produce a health profile. Furthermore, sigmoid activation was used for prediction, and the proposed transformer achieved better accuracy (99.21%), providing a robust framework for ICU patient monitoring.

1 Introduction

In recent years, there has been a rapid increase in healthcare data driven by the widespread adoption of Electronic Health Records (EHRs) and constant monitoring systems in Intensive Care Units (ICUs), which have created novel opportunities for predictive patient monitoring [1]. Precise and timely prediction of adverse outcomes, such as mortality, disease risk, and ICU readmission, helps clinicians make better predictions and enhance patient care [2]. Transformer-based models have emerged as a better solution because of their capability to capture complex temporal dependencies and process heterogeneous multi-modal data such as demographics, laboratory test results, and time-series vital signs. Unlike traditional approaches, which focus on either static or dynamic data in isolation, transformers deliver a unified way to learn both, allowing them to model how clinical events unfold over time and how static patient factors leverage the results. However, the use of Deep Learning (DL) approaches for predictive patient monitoring is not straightforward and has many drawbacks [3]. The first limitation lies in handling incomplete and irregular data, where inconsistencies, noise, and missing values are common [4]. The second limitation is the difficulty of combining heterogeneous types of data, such as demographic information, lab results, and vital signals, into a single representation that is utilized for prediction [5]. Another major drawback is modeling long-term dependencies: patient conditions frequently change steadily and crucial warning signs appear many hours before an event occurs [6]. Moreover, DL approaches also struggle with many issues of interpretability because clinicians not only require accurate prediction, but also perceptions of why the model makes a specific decision [7]. To address these limitations, there is a strong motivation to design advanced DL models that will process multi-modal ICU data in a combined manner and capture both short- and long-term temporal patterns. A Transformer-based predictive system is suitable for this purpose owing to its utilization in self-attention to learn dependencies across time steps and handle large, complex datasets. The aim was to build a model that produced a comprehensive health profile for each patient by integrating demographic details with sequential abs and vitals. We then predicted the mortality risk, disease risk, and readmission with high accuracy. Providing early warning signals, such as systems, has the potential to enhance the workflow of clinical and optimized allocation of resources and ultimately save lives. Several state-of-the-art approaches have been proposed, one of which uses DL approaches on the MIMIC-III dataset to predict patient outcomes [8]. Then, it showed strong performance because it learns from large-scale ICU data automatically without manual feature engineering. However, they fail to capture long-range dependencies and frequently overfit smaller subsets of data. Another approach integrates the static details of patients with sequential data to improve the accuracy of prediction, delivering the advantage of a more holistic modeling of patient health [9]. However, the irregularity and high

* Corresponding author: saef.thallal.iu@gmail.com

variability of ICU time-series signals unable to be fully handled. These drawbacks highlight the requirement of a stronger model that unifies heterogeneous data and effectively models temporal relationships [10]. Against this background, the proposed transformer-based predictive model overcomes these gaps by influencing attention mechanisms, embedding, and positional encoding to produce accurate and interpretable predictions for ICU patient monitoring.

The three main contributions of this research are:

- This research integrates demographic details, laboratory test results and time-series vital signs from the MIMIC-III dataset into a single framework, generating complete view of each patient's health.
- A novel predictive system was designed utilizing transformer architecture with embeddings, positional encoding, and multi-head self-attention, which captures both short-term and long-term patient patterns more efficiently than traditional methods.
- The proposed Transformer-based predictive framework generates highly accurate predictions of the risk of disease, mortality, and ICU readmission, obtaining the highest accuracy and delivering a strong decision-support tool for ICU patient monitoring.

The remainder of this paper is organized as follows. Section 2 provides a literature review, section 3 defines the proposed methodology, section 4 presents the experimental results, and section 5 concludes the research.

2 Literature reviews

Mohapatra et al. [11] illustrated a model for cardiac disease prediction by healthcare big data analysis using an Artificial Neural Network (ANN). Initially, multi-modal health care data were processed by cleaning, normalizing, and encoding by reducing noise from the data. Subsequently, feature extraction and fusion were performed, and attributes were selected and labelled with basic patient details, where they were fed to the ANN model, followed by Particle Swarm Optimization (PSO) and reduced ANN weights for faster computation. Finally, an Artificial Neural Network with Particle Swarm Optimization (ANN-PSO) performed binary classification and achieved superior results. However, this model suffers from different attributes that are limited to categorical data.

Singh et al. [12] presented a predictive model for patient disease risk using the Medical Information Mart for Intensive Care (MIMIC) data. Initially, input data were collected from the MIMIC-III, which consists of the Electronic Health Records (EHR) of Intensive Care Unit (ICU) patients. Preprocessing was performed using data cleaning, one-hot encoding, and normalization. All data were then fed to a gradient boosting regression model to predict the length of stay (LOS), and Long Short-Term Memory (LSTM) was used to predict disease categories. Moreover, the regression model captures complex and non-linear patterns, yet the model struggles with EHR, where incomplete data are available.

Mahbub Ul Alam et al. [13] demonstrated a model named early detection of sepsis from EHR based on the Internet of Medical Things (IoMT). Initially, input data were collected from the MIMIC III dataset and processed using data cleaning, normalization, and Generative Adversarial Imputation Nets (GAIN). Subsequently, the clinical-related features were extracted and fed to the model and then passed to a Recurrent Neural Network with Long Short-Term Memory (RNN-LSTM), which predicts sepsis sequentially from the time-series EHR. Moreover, deploying edge devices using this method may result in low-cost and resource-constrained solutions. However, the demonstrated model struggled with large storage capacity and increased energy consumption.

Zheng et al. [14] proposed a method for predicting diseases using multi-modal graph learning. Initially, the input data were Alzheimer's Disease Prediction of Longitudinal Evolution (TADPOLE) and Autism Brain Imaging Data Exchange (ABIDE). Preprocessing was performed by eliminating unnecessary longitudinal samples and using a Configurable Pipeline for the Analysis of Connectomes (C-PAC) to make the data consistent. Then, a Multi-modal Graph learning (MMGL) model was used to capture the shared and particular required data. Furthermore, Adaptive Graph Learning (AGL) used for automatic learning and to prepare hidden graph structures from data. Finally, robust predictions were performed using a Graph Convolution Network (GCN), but the model still suffers from large datasets and requires long computation time.

Siu et al. [15] demonstrated a health guardian using multi-modal data to understand individual health. Initially, input data were collected from a long-term movement-monitoring database (LTMM). Next, the data were fed for pre-processing, which included de-identifying the health information, imputing the missing values, integrating timestamps from various sensors, and making the data consistent. Features were extracted using the Patient Health Questionnaire (PHQ-8) and Amyotrophic Lateral Sclerosis (ALS). Finally, the preprocessed data were fed to the Random Forest (RF) model to analyze and predict the wearable sensor mobility. Furthermore, the model performs early detection and provides broad access to health care assessments. However, the proposed methods are sensitive to environmental factors that affect stability and reliability.

3 Methodology

The proposed transformer-based predictive framework follows a standard workflow to observe patients in care Units (ICU) patients and predict crucial outcomes. First, data were collected from the Medical Information Mart for Intensive Care III (MIMIC-III) dataset, which consists of demographic details, laboratory test results, and time-series vital signs. Next, the data were cleaned to remove duplicates, normalized to scale the values within a uniform range, and segmented to divide each

patient's ICU stay into fixed-length time windows, ensuring that the data were appropriate for sequential analysis. The segmented data are then passed to the transformer model, and each time step is transformed into embedding and positional encoding to retain the order of events. Furthermore, data are processed through a multi-head self-attention mechanism that captures essential temporal patterns and highlights crucial clinical signals. Additionally, feedforward layers integrated information into a compact health profile, which was then passed to an output layer with sigmoid activation to predict the probability of three probabilities for disease risk, mortality risk, and ICU readmission. Figure 1 shows the overall workflow of the proposed transformer-based predictive framework.

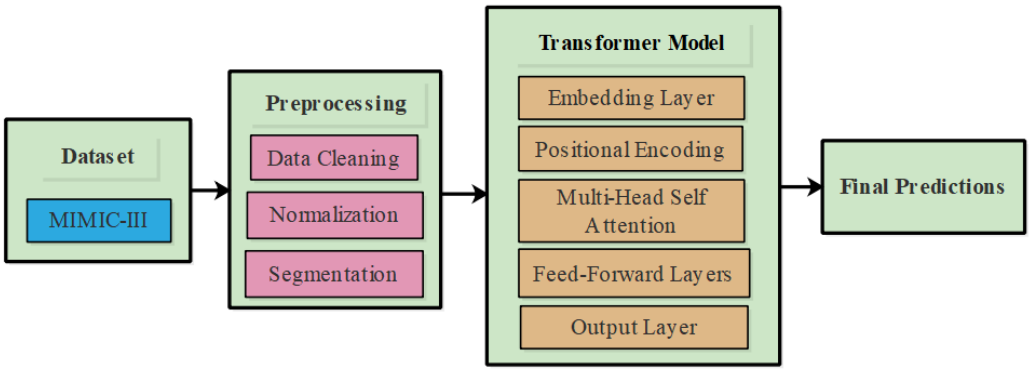

Fig. 1. Overview of proposed Transformer-based predictive framework.

3.1 Data collection

In this research, data were collected from the MIMIC-III database, which is one of the most broadly used critical care datasets for the prediction of healthcare research. The database consists of detailed Electronic Health Records (EHRs) of more than 58,000 ICU admissions between 2001 and 2012 [12]. All records were fully anonymized to ensure patient privacy while providing rich clinical information. The dataset consisted of three major types of information: demographic and admission details, including sex, age, type of admission, length of stay, marital status, ethnicity, and insurance. Second, laboratory test results that record measurements, such as glucose, creatinine, cholesterol, hemoglobin, and other biomarkers, provide clinical insight into patient health. Thirdly, time-series vital signs that includes blood pressure, heart rate, oxygen saturation, body temperature and respiratory rate which are recorded at many time points at the time of patients ICU stay. Together, these provide a combination of static and dynamic features that are beneficial for monitoring patient condition. For this research, relevant data were extracted from many MIMIC-III tables, such as patients, admissions, ICU stays, lab events, and chart events. Records with missing or invalid values were eliminated to ensure the quality. The final dataset integrated demographic information, laboratory measurements, and time-series vital signs to form a comprehensive foundation for training the proposed transformer-based predictive model for disease prediction, mortality risk, and ICU readmission.

3.2 Preprocessing

The raw dataset collected from MIMIC III contains patient demographics, laboratory test results, and time-series vital signs. However, such medical records are not directly used for modeling, because they frequently include noise, redundancy, inconsistent formats, and missing entries. Preprocessing techniques were applied to confirm that the dataset was clean, standard, and structured before it passed into the transformer model. Three main preprocessing techniques were applied: data cleaning, normalization, and segmentation.

3.2.1 Data cleaning

The raw dataset consisted of missing values, duplicate entries, and inconsistencies in medical records. Initially, duplicate records were identified and removed to avoid redundancy, and demographic details, such as age, sex, and type of admission, were handled by eliminating incomplete patient entries, ensuring that only complete demographic records were retained. Third, statistical techniques such as mean and median imputation are used to handle missing values in laboratory results and time-series vital signs. Next, linear interpolation was used to estimate the values that ensured continuity in patient records. Finally, invalid values and outliers were detected, corrected to within an acceptable range, or eliminated. This data-cleaning step ensured that all three modalities (demographics, laboratory tests, and vital signs) were consistent, accurate, and ready for further processing. Mathematically, the data-cleaning process is expressed by Equation (1).

$$D_{clean} = f_{dup}(f_{miss}(f_{out}(D_{raw}))) \quad (1)$$

Here, f_{dup} denotes duplicate removal, f_{miss} denotes imputation for missing values, f_{out} denotes corrections of outliers, D_{raw} denotes the raw dataset, and D_{clean} denotes the final cleaned dataset. Therefore, a clean dataset with scalable demographic details, consistent laboratory test values, and valid vital signs was obtained and ready for further processing.

3.2.2 Normalization

Even after cleaning, the dataset contained numerical features with very different scales to ensure uniformity, and min-max normalization was applied. The min-max normalization scales all the numerical values in the range of 0 to 1, which prevents large-value features from dominating the smaller values. This improves model stability, and the process is mathematically expressed by Equation (2).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

Here, x signifies the original value, x_{min} signifies the minimum value, x_{max} signifies the maximum value, and x' signifies the normalized value. Hence, a normalized dataset was obtained, in which all numerical values were scaled consistently between 0 and 1.

3.2.3 Sliding window segmentation

The ICU records collected from the dataset are continuous and vary in length, making them inappropriate for direct input to the model. To overcome this challenge, each patient's ICU data was divided into fixed-length time windows using sliding-window segmentation. The time-series vitals and repeated lab results were split into windows, whereas static demographic features were associated with every segment in the patient context. These segments ensure that each input represents a constant period of patient history and are formally defined in Equation (3):

$$S = f_{seg}(D_{norm}, W) \quad (3)$$

where f_{seg} represents the operation that divides the continuous ICU time-series data into fixed-length segments, D_{norm} represents the normalized dataset, W is the fixed window size, and S is the final segmented data. Finally, a structured dataset of patient segments was obtained and was ready to be fed into the transformer model.

3.3 Transformer-based predictive framework

After pre-processing, the segmented ICU dataset was forwarded to a transformer-based predictive framework. The framework processes static and dynamic features to learn complex temporal patterns to predict the risk of disease, mortality, and readmission. This workflow is explained in detail in the following section.

3.3.1 Embedding layer

A segmented dataset consisting of normalized lab values, time-series vitals, and static demographics is not directly fed to the transformer because it requires dense vector representations. Therefore, an embedding layer was used to transform each feature at every time step into dense embedding. This allows all patient features to be presented in a standard format appropriate for the model, and is statistically expressed in Equation (4).

$$E_t = f_{embed}(x_t) = W_e . x_t \quad (4)$$

Here, W_e illustrate the embedding weight matrix, x_t the feature vector at time step t and f_{embed} the embedding operation. Therefore, dense patient embeddings were obtained from this step, demonstrating each time step as a fixed-length vector.

3.3.2 Positional encoding

The patient embeddings from the embedding layer lacked information regarding the order of the events; therefore, the model was unable to identify which event occurred earlier or later. Then, positional encoding is added to each embedding by utilizing sine and cosine functions of different frequencies. This provides the temporal order of the data, which is mathematically expressed by Equations (5)-(6).

$$PE_{(pos,2i)} = sin\left(\frac{pos}{10000^{2i/d}}\right) \quad (5)$$

$$PE_{(pos,2i+1)} = cos\left(\frac{pos}{10000^{2i/d}}\right) \quad (6)$$

Here, pos is the position of the time step, i is the index of the dimension, d is the embedding dimension, and position-aware embedding is obtained, including patient features and their sequence order.

3.3.3 Multi-head self-attention

Traditional approaches struggle with long-term dependencies across patient records; therefore, the self-attention mechanism compares each time step in a sequence with all other time steps. It assigns higher weights to essential events such as sudden spikes in blood pressure or drops in oxygen levels. In parallel, multiple attention heads operate, enabling the model to capture the various types of dependencies. Here, static demographic features are not processed but are later concatenated with the attention output for context. Mathematically, the multihead self-attention is illustrated in Equation (7).

$$MHA(X) = Concat(head_1, head_2, ..., head_n)W_O \quad (7)$$

Here, W_O denotes the final projection matrix, X denotes positional-aware embeddings, and finally, context-rich sequence representations are obtained, where essential health patterns in time-series vitals and laboratory results are highlighted and later integrated with static features in the next step.

3.3.4 Feed-forward layers

Although self-attention provides a powerful context, it produces high dimensionality, which is difficult to interpret directly. They include several parallel features from multiple heads that must be integrated into a more compact form. Therefore, each output vector from self-attention passes to a position-wise feed-forward network

that involves two fully connected layers with nonlinear activation between them. This compresses the information and highlights the most essential clinical features. Mathematically, this is expressed by Equation (8).

$$FFN(x) = max(0, xW_1 + b_1)W_2 + b_2 \quad (8)$$

where W_1, W_2 are learnable weight matrices, b_1 *and* b_2 are bias terms, and $max(0,.)$ is ReLU activation. In this step, the static demographic features are concatenated with the compressed results to enrich the profile with static and dynamic information. Therefore, the comprehensive health profile was compact, discriminative, and ready for the final prediction step.

3.3.5 Output layer

The comprehensive health profile vector generated by the feedforward layers integrates the temporal patterns from the vitals with static demographic details. Here, the health profile contains rich patient information and has continuous feature values that are not directly understandable for clinical decision-making. To make the predictions useful, these values must be transformed into probabilities that reflect the likelihood of a particular outcome. A Sigmoid activation function was applied to transform the scores to probabilities in the range of 0 to 1, enabling them to be interpretable for prediction, this is expressed by Equation (9).

$$P(y = 1|x) = \sigma(z) = \frac{1}{1+e^{-z}} \quad (9)$$

Where $\sigma(z)$ signifies the probability that outcome y occurs, $P(y = 1|x)$ signifies the predicted probability given patient input x, z signifies the output score. Therefore, the final prediction probabilities for the three outcomes (disease risk, mortality risk, and ICU readmission) were calculated.

4 Experimental results

In this research, the proposed transformer-based predictive framework for ICU patient monitoring was implemented using Python 3.9, TensorFlow 2.12, and PyTorch 1.13. Pandas and NumPy were used in the preprocessing pipeline, whereas Scikit-learn was used for the normalization and segmentation processes. The model was trained on an NVIDIA Tesla V100 GPU. Evaluation metrics such as accuracy, precision, recall and F1-score. These metrics are expressed by Equations (10)–(13) as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

$$F1 - Score = \frac{2.Precision.Recall}{Precision+Recall} \quad (13)$$

Here, TP is denoted as True Positive, TN as True Negative, FP as False Positive, FN as False Negative.

4.1 Performance analysis

The performance of the proposed transformer-based framework was compared with two traditional approaches: RNN-LSTM and Graph Convolutional Network (GCN). This evaluation was performed using the accuracy, precision, recall and F1-Score. The performance analysis of the proposed system is presented in Table 1.

Table 1. Performance analysis.

Performance Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RNN-LSTM	96.42	96.05	95.80	95.92
GCN	97.10	96.82	96.55	96.68
Proposed Transformer	99.21	99.10	99.15	99.12

Table 1 demonstrates the superior performance of the proposed transformer-based predictive framework in contrast to traditional approaches. The proposed transformer achieved significant improvements in accuracy (99.21%), precision (99.10%), recall (99.15%), and F1-score (99.12%), highlighting the scalability and robustness of ICU patient monitoring.

4.2 Comparative analysis

The efficiency of the proposed transformer-based predictive framework was further validated by performing a comparative analysis against the existing ANN+PSO [11] method. A comparative analysis of the proposed transformer-based predictive framework is presented in Table 2.

Table 2. Comparative analysis.

Comparative Model	Accuracy (%)	Precision (%)	Recall (%)
ANN+PSO [11]	99.02	98.88	98.92
Proposed Transformer	99.21	99.10	99.15

Table 2 highlights that the proposed Transformer-based framework constantly surpasses the existing ANN+PSO [11] approach in terms of all key metrics such as accuracy (99.21%), precision (99.10%) and recall (99.15%). The transformer effectively combines demographics, time-series vitals and self-attention mechanism which ensures accurate temporal modeling. Hence, the proposed Transformer-based framework is more robust and scalable for real-time ICU patient monitoring.

4.3 Discussion

In this research, a transformer-based predictive framework was proposed for ICU patient monitoring with the objective of predicting disease risk, mortality, and readmission. The existing ANN-PSO achieved a high classification accuracy but faced significant challenges. One major challenge is its inability to capture complex temporal dependencies in ICU time-series data and its restricted capacity to integrate static patient data with dynamic data streams such as demographics. This limits the overall contextual understanding, and the constraints reduce the adaptability and robustness of the model in a real clinical environment. The proposed transformer-based predictive framework addresses these drawbacks by introducing embeddings and positional encodings that preserve the sequential order of patient events, while confirming the uniform illustration of both static and dynamic features. Furthermore, the multi-head self-attention mechanism allows the model to focus on the most critical clinical updates across time windows, allowing sudden deterioration to be overlooked. Moreover, the feed-forward layers condense these perceptions into a comprehensive health profile for a robust prediction. Finally, the proposed transformer-based predictive framework illustrates superior predictive performance and ensures more scalable, real-time support for ICU clinicians compared with the existing ANN-PSO approach.

5 Conclusion

The primary objective of this research was to design an effective and scalable framework for predictive patient monitoring in ICUs that accurately predicts disease risk, mortality, and readmission using a transformer-based predictive framework. Initially, data, including patient demographics, laboratory results, and vital signs, were collected from the MIMIC III dataset. Preprocessing was performed by cleaning, normalization, and segmentation. The preprocessed data are then fed to the transformer model, where embeddings and positional encodings illustrate the patient features and preserved temporal order. Furthermore, a multi-head self-attention mechanism is used to capture essential dependencies through time steps, and feed-forward layers produce a compact health profile for each patient. Sigmoid activation was applied to generate probabilities of disease, mortality, and readmission. Therefore, the proposed transformer-based predictive framework obtained better performance across all metrics: accuracy (99.21%), precision (99.10%), and recall (99.15%), compared with the existing ANN-PSO approach, and generated a robust solution for ICU patient monitoring. In future, transformer-based predictive framework will be extended by combining additional data sources like medical imaging and clinical notes, and deploying it in real-time ICU settings.

References

1. G. Ramkumar, J. Seetha, R. Priyadarshini, M. Gopila, G. Saranya, IoT-based patient monitoring system for predicting heart disease using deep learning. Meas. **218**, 113235 (2023). https://doi.org/10.1016/j.measurement.2023.113235
2. N. Alharbe, M. Almalki, IoT-enabled healthcare transformation leveraging deep learning for advanced patient monitoring and diagnosis. Multimed. Tools Appl. **84**, 21331-21344 (2025). https://doi.org/10.1007/s11042-024-19919-w
3. A. Motwani, P.K. Shukla, M. Pawar, Novel framework based on deep learning and cloud analytics for smart patient monitoring and recommendation (SPMR). J. Ambient Intell. Humaniz. Comput. **14**, 5565-5580 (2023). https://doi.org/10.1007/s12652-020-02790-6
4. H. Ni, S. Meng, X. Geng, P. Li, Z. Li, X. Chen, X. Wang, S. Zhang, Time series modeling for heart rate prediction: From arima to transformers, In 2024 6th International Conference on Electronic Engineering and Informatics (EEI), IEEE, Chongqing, China, October 08 (2024), 584-589
5. S. Aghalya, N.V. Rao, R.R. Roy, C. Srinivasan, G. Premi, Real-time Monitoring and Prediction of Respiratory Diseases Using IoT and Machine Learning. In 2023 Second International Conference on Smart Technologies for Smart Nation (SmartTechCon), IEEE, Singapore, Singapore, January 19 (2024), 635-640
6. S.P. Vimal, M. Vadivel, V.V. Baskar, V.G. Sivakumar, C. Srinivasan, Integrating IoT and machine learning for real-time patient health monitoring with sensor networks, In 2023 4th International Conference on Smart Electronics and Communication (ICOSEC), IEEE, Trichy, India, October 16 (2023), 574-578
7. J.A.J. Singh, J. Gnanasoundharam, M. Birunda, G. Sudha, S.P. Maniraj, C. Srinivasan, Wearable Sepsis Early Warning Using Cloud Computing and Logistic Regression Predictive Analytics, In 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), IEEE, Noida, India, May 14 (2024), 1-6
8. A. Cuevas-Chávez, Y. Hernandez, J. Ortiz-Hernandez, E. Sanchez-Jimenez, G. Ochoa-Ruiz, J. Perez, G. Gonzalez-Serna, A systematic review of machine learning and IoT applied to the prediction and monitoring of cardiovascular diseases. Healthcare **11,** 2240 (2023). https://doi.org/10.3390/healthcare11162240
9. A. Ahmad, H.K. Hussain, H. Tanveer, T. Kiruthiga, K. Gupta, The intelligent heart rate monitoring model for survivability prediction of cardiac arrest patients using deep cardiac learning model, In 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), IEEE, Coimbatore, India, April 19 (2023), 376-381

10. K.K. Baseer, K. Sivakumar, D. Veeraiah, G. Chhabra, P.K. Lakineni, M.J. Pasha, R. Gandikota, G. Harikrishnan, Healthcare diagnostics with an adaptive deep learning model integrated with the Internet of medical Things (IoMT) for predicting heart disease. Biomed. Signal Process. Control **92**, 105988 (2024). https://doi.org/10.1016/j.bspc.2023.105988
11. S. Mohapatra, P.K. Sahoo, S.K. Mohapatra, Healthcare big data analysis with artificial neural network for cardiac disease prediction. Electron. **13**, 163 (2023). https://doi.org/10.3390/electronics13010163
12. D. Singh, V. Kumar, R.G. Qiu, Patients' disease risk predictive modeling using MIMIC data. Procedia Comput. Sci. **168**, 112-117 (2020). https://doi.org/10.1016/j.procs.2020.02.256
13. M.U. Alam, R. Rahmani, Fedsepsis: A federated multi-modal deep learning-based internet of medical things application for early detection of sepsis from electronic health records using raspberry pi and jetson nano devices. Sensors **23**, 970 (2023). https://doi.org/10.3390/s23020970
14. S. Zheng, Z. Zhu, Z. Liu, Z. Guo, Y. Liu, Y. Yang, Y. Zhao, Multi-modal graph learning for disease prediction. IEEE Trans. Med. Imaging **41**, 2207-2216 (2022). https://doi.org/10.1109/TMI.2022.3168783
15. V.S. Siu, K.Y. Hsieh, I. Buleje, T. Itoh, T. Hao, B. Civjan, N. Hinds, B. Dang, J.L. Rogers, B. Wen, Health Guardian: Using Multi-modal Data to Understand Individual Health, In 2023 IEEE International Conference on Digital Health (ICDH), IEEE, Chicago, IL, USA, August 28 (2023), 65-74