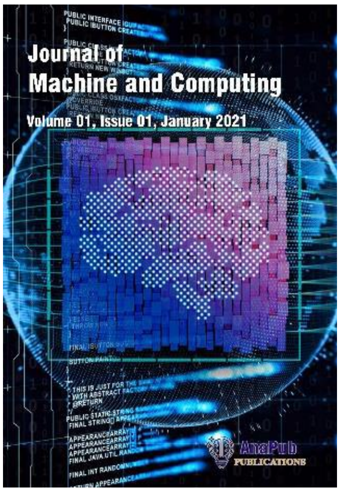
Journal Pre-proof



Meta-Learning for Enhanced Disease Prediction from EHR Data

Chitra G and Syed Muzamil Basha

DOI: 10.53759/7669/jmc202505169
Reference: JMC202505169
Journal: Journal of Machine and Computing.

Received 06 April 2025
Revised from 22 June 2025
Accepted 21 July 2025

Please cite this article as: Chitra G and Syed Muzamil Basha, "Meta-Learning for Enhanced Disease Prediction from EHR Data", Journal of Machine and Computing. (2025). Doi: https:// doi.org/10.53759/7669/jmc202505169.

AnaPub
PUBLICATIONS

Meta-Learning for Enhanced Disease Prediction from EHR Data

Chitra G [1] and Syed Muzamil Basha[1]

[1]School of Computer Science and Engineering, REVA University, Bengaluru, India,
lakki.gr@gmail.com, muzamilbasha.s@reva.edu.in

***Abstract*—**The healthcare sector is becoming more dependent on electronic health records (EHR) for disease forecasting, risk evaluation, and mortality analysis. Although AI-driven models have enhanced disease prediction, they frequently focus on common diseases and face difficulties with new or rare diseases. Furthermore, these models require large datasets for better accuracy, posing challenges in diverse or limited-data scenarios. To solve these issues, this research proposes a novel Long Short-Term Memory (LSTM)-Attention network-based meta-learning framework for prediction tasks using time-series data from EHRs. The framework is designed to address challenges such as limited sample sizes, imbalanced labels, and the ability to predict unseen diseases. The proposed model is capable of handling multiple tasks related to irregular patterns and anomalies in time-series signals. The meta-learning approach enables the system to leverage knowledge from previous tasks, enhancing its ability to predict new and previously unseen diseases from ECG data. The proposed LSTM-Attention model is evaluated against conventional models like Support Vector Machine (SVM), Random Forest (RF), and XGBoost. Experimental results demonstrate that the proposed model outperforms these models, achieving superior performance in predicting HRV, arrhythmia, and abnormalities from ECG signals. The LSTM-Attention model achieves the highest accuracy (0.92), precision (0.90), recall (0.91), F1 score (0.91), and ROC-AUC (0.93). Moreover, the prediction time for the proposed model is 95 seconds, significantly faster than other models.

Keywords*—Electronic Health Record; Meta-Learning; Electrocardiogram; LSTM with Attention; Disease Prediction; Accuracy.*

I. INTRODUCTION

Continuous patient Electronic Health Records (EHR) have received a lot of attention in healthcare analytics recently [1]. Many studies have been conducted to create strategies for predicting clinical risks such as death, hospital readmission, the onset of a chronic disease, and deterioration of an existing condition. The reasons for this include: 1) the difficulty of analyzing patient EHRs due to factors such as noise, sparsity, and inconsistency, and 2) the need for reliable health risk prediction models to assist clinical decision-makers in identifying potential risks early on so that patients can receive better care. In response to these challenges, a variety of computer algorithms have been developed, ranging from more standard methods to Deep Learning (DL) models [2].

A significant difference between healthcare challenges and applications in domains such as robot vision, speech evaluation, and machine translation is the scarcity of available sample datasets, as well as the high cost or inability to collect additional samples [3]. Each data sample is paired with a specific patient in the context of personalized patient risk prediction, which aims to forecast a specific clinical risk for each individual. With a narrower focus on a certain condition, the already small world population of 7.5 billion people decreases even further. Unfortunately, there is only a limited quantity of patient data available in a certain EHR corpus for training a risk prediction model. Furthermore, the clinical dangers being addressed are extremely detailed. A lack of comprehensive understanding of the biochemical pathways underlying the majority of deadly diseases complicates the design of effective treatments. To develop models that can reliably predict clinical risk, it is essential to maximize the use of limited patient samples alongside existing knowledge about clinical risk and predictive models. This study specifically focuses on predicting illness using ECG. ECGs provide a wealth of critical physiological information for diagnosing a wide range of cardiac conditions. Their intricate patterns carry vital insights, yet understanding these signals often requires highly trained specialists; even experienced doctors may struggle to identify subtle trends or apply their knowledge to new or challenging patients.

A. *Current Challenges and Proposed Solution*

Recognizing the complexity of ECG analysis underscores the need for advanced computational techniques to streamline and enhance diagnostic accuracy. The traditional Machine Learning (ML) pipeline for illness diagnosis involves two main steps: feature extraction and model development. Initial methods primarily focused on using a single characteristic for diagnosis. However, these standard approaches have proven insufficient due to the high level of manual involvement and the requirement for specialized expertise. To address these limitations, DL, particularly convolutional neural networks (CNN), have become powerful alternatives for disease prediction. These networks automate feature extraction and optimize performance, thus minimizing manual intervention in parameter selection. However, DL models require large datasets for accurate predictions and are often applicable only for identifying specific diseases. They are not suitable for predicting rare diseases or conditions with minimal samples.

To address these challenges, a meta-learning approach is proposed in this work. The LSTM-Attention model is chosen as the base model for meta-learning. The proposed system analyzes ECG data and is capable of predicting HRV, arrhythmia, and other abnormalities in ECG readings. This model is beneficial for predicting rare diseases from ECG signals, thereby improving patient health and facilitating appropriate treatment.

B. *Problem Statement and Research Contribution*

Identifying diseases from ECG requires more data and is affected by data quality issues such as noise, sparsity, and inconsistency. Traditional methods fail to identify rare diseases due to limited datasets. This paper proposes a solution that overcomes these challenges by using meta-learning combined with an LSTM-Attention model. The key contribution of the research:

- The meta-learning approach is utilized to predict HRV, arrhythmia, and other abnormalities using a single model from minimal ECG data.
- An LSTM-Attention DL model is proposed to capture temporal features from ECG data, with attention mechanisms incorporated to identify the most significant features for enhanced prediction accuracy.
- The proposed LSTM-Attention model is compared with traditional models such as SVM, XGBoost, and RF using quality metrics and execution time.
- The EHR data is sourced from MIMIC-III, which is highly unstructured and underwent extensive pre-processing to enhance the accuracy of DL-based meta-learning prediction.

The research paper is organized as follows: Section I details the importance of disease prediction from EHR, challenges in current methods, and the need for meta-learning. Section II discusses recent research work on EHR data. Section III covers the theoretical concept of meta-learning and the LSTM-Attention model. Section IV discusses the experimental setup, data used, and outcomes of the proposed model in disease prediction. Section V concludes the research with future work.

II. Related Work

Various risk prediction models have been developed using ML, DL, and EHR data. Some of the recent works in this area are highlighted in this section. Using feature extraction, the study [4] proposed an EHR risk prediction model. By initially extracting structured text from EHR data using Natural Language Processing, the researchers [5] recommend a system that employs ML techniques to categorize the text as an indicator of "good" or "bad" quality and then use it for prediction. Logistic regression and SVM were utilized as ML models. The research [6] proposed a system consisting of two models. The initial model included the development of interoperable EHR systems with a uniform database structure. Module two covered tasks including cleaning and retrieving data from the EHR system, as well as evaluating and forecasting data. Using the proposed decision support system, the proposed system's accuracy in forecasting diabetes disease and the EHR system's interoperability were assessed.

The study [7] developed a multi-task learning (MTL) model to make clinical predictions from time-series EHR data. To demonstrate that MTL systems can overcome task imbalance and interference, the study [8] compares their performance to that of traditional single-task models using a prenatal EHR dataset. With uniform input durations and variables, these models may be applied to all patients in the EHR system. Transfer learning was employed in one study [9] to address the diminishing data problem in personalized models. The study [10] aimed to increase disease prediction accuracy by utilizing EHR temporal data through a new hybrid DL architecture. The architecture incorporates both CNN and LSTM networks. Findings support the notion that predictive model development should shift toward including complex neural network topologies, potentially leading to more personalized models. The paper [11] proposes a model that integrates LSTM and Graph Neural Networks (GNNs) to forecast opioid overdose risks, utilizing temporal illness development and patient interaction graphs. The model's dependence on extensive EHR data may raise privacy issues, and although interpretability has improved, it may still provide difficulties for doctors in real-time scenarios. The research [12] examines DL techniques for clinical decision support, utilizing electronic health record data to predict illness stages, detect genetic markers, and anticipate hospitalization requirements. Although DL provides great accuracy, it remains constrained by the availability and quality of labeled data, as well as the complexity of integrating various electronic health record formats.

The study [13] employs LSTM to evaluate structured EHR data for the automated prediction of surgical site infections, surpassing conventional models such as random forests in both precision and area under the ROC curve. Although LSTM models exhibit excellent accuracy, they necessitate extensive, pristine, and meticulously annotated datasets, which can be difficult to acquire consistently across various healthcare environments. The research [14] proposes a hybrid model that integrates LSTM and machine learning techniques to predict new-onset delirium based on patient data, surpassing the performance of conventional models such as logistic regression and LightGBM. The efficacy is significantly reliant on the quality of EHR data and feature selection, and the model may encounter difficulties with infrequent or unrecognized conditions. Paper [15] suggests extracting and predicting lung cancer from EHR datasets using a DL architecture combined with Natural Language Processing (NLP). The text mining model can automatically forecast occurrences from input datasets, allowing for optimal cancer predictions. The model was tested in a new context to see how well it performs and how resilient DL with NLP is to different datasets. The evaluation demonstrated that the prediction accuracy was higher than that of existing methods. However, there are some limitations in the DL models. The scarcity of labeled datasets and datasets with long-tailed class distributions complicates a wide range of medical tasks. Because general practitioners may be unfamiliar with rare diseases and struggle to differentiate between them, an AI-based decision-making model could help improve diagnostic accuracy.In recent research, many algorithms have been proposed for handling EHR data and predicting diseases. However, there remains a gap between research and real-time implementation. Proposed models often struggle with unseen data, rare diseases, and new conditions. Most studies use data collected from the internet, which is typically sourced from specific groups of people within certain locations or age ranges. This limitation makes such data insufficient for deploying models in clinical settings. The research aims to address these problems using an LSTM-Attention-based meta-learning approach.

III. THEORETICAL BACKGROUND

This section details the theoretical foundation of two important concepts in this research: meta-learning and LSTM-Attention networks. It also explains how to integrate the LSTM-Attention model into the meta-learning framework. The nomenclature used in the meta-learning and LSTM-Attention networks is provided in Appendix 1 for reference.

A. Meta-Learning

In this case, the model is designed to learn from a limited number of samples. The key concept behind meta-learning is to build on prior knowledge rather than starting from scratch to achieve the goal [6]. This approach draws upon foundational concepts from supervised machine learning. Here is a model that accepts an observation x and assigns it a label y. D represents the training data, while ϕ represents the model parameters. Training is an optimization challenge in supervised learning, to maximize the likelihood of the parameters based on the D : $\arg max_{\phi} \log p(\phi|D)$. The issue is reframed as the determinant of the limit of the parameter's marginal probability: $\arg max_{\phi} \log p(\phi)$and the data's maximum probability given the parameters: $\arg max_{\phi} \log p(D|\phi)$. A regularizer, such as $\log p(\phi)$ (e.g., weight decay), can be used, and optimization is performed over the dataset with $\sum_i \log p(y_i|x_i,\phi)$. Equations (1-4) describe supervised learning for optimizing model parameters based on task-specific data.

$$\arg max_{\phi} \log p(\phi|D) = \arg max_{\phi} log \frac{p(\phi,D)}{p(D)} \tag{1}$$

$$= \arg max_{\phi} \log p(\phi|D) \tag{2}$$

$$= \arg max_{\phi} \log p(D|\phi) + \log p(\phi) \tag{3}$$

$$= \arg max_{\phi} \textstyle\sum_i \log p(y_i|x_i,\phi) + \log p(\phi) \tag{4}$$

Given $D = \{(x_1, y_1), \ldots, (x_k, y_k)\}$ where x_i denotes the input and y_i denotes the corresponding labels, meta-learning integrates prior experiences with limited new data inputs. To merge existing knowledge with new data points, the problem is expressed in Equation (5)

$$\arg max_{\phi} \log p(\phi|D, D_{meta-train}) \tag{5}$$

Where $D_{meta-train}$ represents prior or meta-training data. Maintaining $D_{meta-train}$ is difficult due to significant memory demands. Meta-learning characterizes the meta-training dataset through meta-parameters obtained by $\theta = \arg max_{\theta} p(\theta|D_{meta-train})$. The meta-parameters θ, derived from $D_{meta-train}$, encapsulate prior information essential for the rapid execution of new tasks. This task becomes a maximum likelihood issue: $\arg max_{\phi} \log p(\phi|D, D_{meta-train})$. Using prior meta-training data, the goal is to optimize the likelihood of the parameters concerning the new data. The likelihood process is regarded as an integration of the θ. To estimate this integration, a point estimate θ^* for the $\theta$is used. Meta-training $p(\theta^*|D_{meta-train})$ involves acquiring meta-parameters from existing meta-training data, while adaptation $p(\phi|D,\theta^*)$ focuses on deriving parameters for a novel task using both new data and established meta-parameters. Equations (6-10) describe meta-learning for optimizing model parameters using prior knowledge from meta-training data to enable fast adaptation to new tasks.

$$log\ p(\phi|D, D_{meta-train}) = log \frac{p(\phi,D,D_{meta-train})}{p(D,D_{meta-train})} \tag{6}$$

$$= log \int \frac{p(\phi,D)p(\theta,D_{meta-train})}{p(D,D_{meta-train})} d\theta \quad (7)$$

$$= log \int \frac{p(\phi,D,\theta)p(\theta,D_{meta-train})}{p(D,\theta)p(D_{meta-train})} d\theta \quad (8)$$

$$= log \int p(\phi|D,\theta)\, p(\theta|D_{meta-train})\, d\theta \quad (9)$$

$$\approx \log p(\phi|D,\theta^*)p(\theta^*|D_{meta-train}) \quad (10)$$

Where $\theta^* = \arg max_\theta\, p(\theta|D_{meta-train})$. To estimate the new task $\arg max_\phi \log p(\phi|D, D_{meta-train})$, use $\arg max_\phi \log p(\phi|D,\theta^*)$, where ϕ represents the task-specific parameters and θ^* denotes the initial information distributed across all tasks.

In summary, acquiring a new skill involves two stages. The first stage is gaining proficiency with the meta-learning parameter $\theta^* = \arg max_\theta\, p(\theta|D_{meta-train})$, followed by refining it using a limited number of examples $\phi^* = \arg max_\phi\, p(\phi|D,\theta^*)$.

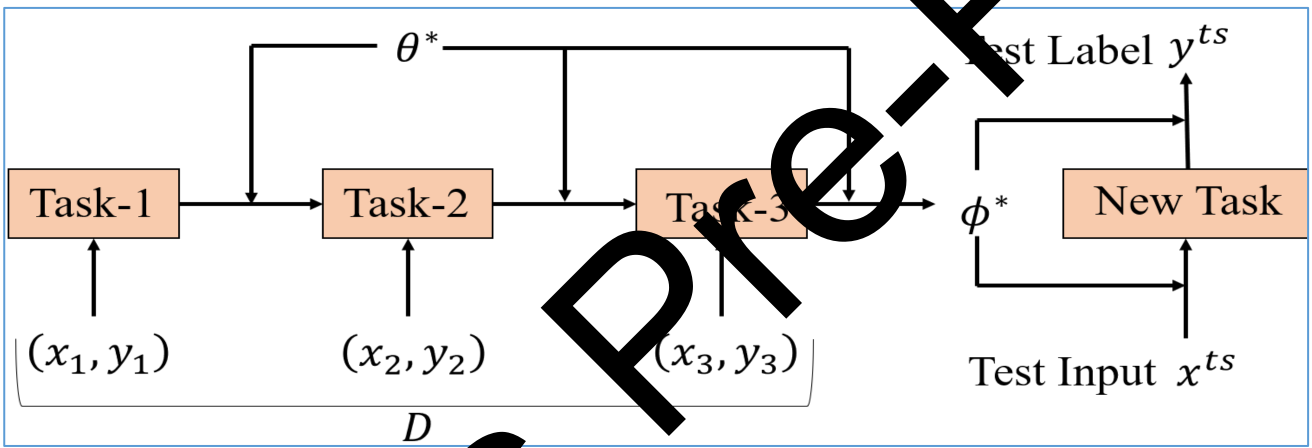


Fig. 1. Meta-Learning for new task prediction.

Figure 1 shows the working of meta-learning in new task prediction and it illustrates the involvement of meta-parameters in new tasks and how meta-parameters are updated based on task-specific parameters using a limited number of new input/output pairs. Ultimately, the updated ϕ^* is utilized for prediction. The algorithm 1 gives the working of meta-learning model.

Algorithm 1: Meta-Learning

Input: Meta-training data $\boldsymbol{D_{meta-train}}$, New task data $\boldsymbol{D_{new}}$, Parameters $\boldsymbol{\theta}, \boldsymbol{\phi}$

Output: Adapted model parameters $\boldsymbol{\phi}^*$, performance.

Step 1: Initialize Meta-parameters $\boldsymbol{\theta}$.

For each task $\boldsymbol{T_i}$ in $\boldsymbol{D_{meta-train}}$:

a. Extract the training data $\boldsymbol{D_i}$ for task $\boldsymbol{T_i}$.
b. Train the model on task $\boldsymbol{T_i}$ using the current $\boldsymbol{\theta}$.
c. Update meta-parameters $\boldsymbol{\theta}$.

$$\boldsymbol{\theta^* = \arg max_\theta\, p(\theta|D_{meta-train})}$$

Step 2: Adapt at New Task

a. Extract New Task Data
b. Initialize Task-Specific Parameters $\boldsymbol{\phi}$

---

c. Update task-specific parameters $\boldsymbol{\phi}$ using new task data $\boldsymbol{D}$ and the prior $\boldsymbol{\theta}^*$.

$$\boldsymbol{\phi}^* = \mathbf{arg}\,\boldsymbol{max}_{\boldsymbol{\phi}}\, \boldsymbol{p}(\boldsymbol{\phi}|\boldsymbol{D}, \boldsymbol{\theta}^*)$$

Step 3: Evaluate the Adapted Model

Step 4: Repeat Steps 2 and 3 to adapt the model to different tasks, using the prior $\boldsymbol{\theta}^*$.

B. *LSTM-Attention*

This section details the importance of combining LSTM and the Attention mechanism. First, the architecture and working of LSTM and the Attention mechanism are detailed. Next, the proposed model is explained.

LSTM: When handling large volumes of sequencing data, LSTM is recommended as a training method [[illegible]7]. The gated state enables LSTM to exercise more control over the transmission state compared to the standard RNN. After a long state $c_{(t-1)}$, the forget gate $f_{(t)}$ retains or discards the input. The input gate $c_{(t-1)}$ determines whether to allow the input data to the current state of the memory cell $c_{(t)}$. The output gate $o_{(t)}$ functions similarly to the input gate by deciding to send the signal from the current to the next layer. Equations (11-15) illustrate the calculation principles for each gate structure [24].

$$f_{(t)} = \sigma\left(W_f x_{(t)} + V_f h_{(t-1)} + b_f\right) \tag{11}$$

$$i_{(t)} = \sigma\left(W_i x_{(t)} + V_i h_{(t-1)} + b_i\right) \tag{12}$$

$$o_{(t)} = \sigma\left(W_o x_{(t)} + V_o h_{(t-1)} + b_o\right) \tag{13}$$

$$c_{(t)} = f_t \otimes c_{(t-1)} + i_{(t)} \otimes tanh\left(W_c x_{(t)} + V_c h_{(t-1)} + b_c\right) \tag{14}$$

$$h_{(t)} = o_{(t)} \otimes tanh\left(c_{(t)}\right) \tag{15}$$

The input sequence $x = (x_1, x_2 \ldots x_n)$ and the output sequence $h = (h_1, h_2 \ldots h_n)$ are defined as follows. The weight matrix W and vector V re-established, with b representing the threshold. The sigmoid activation function $\sigma(x) = {}^{1}\!/_{1\,+\,e^{-x}}$ is defined, and the dot product is indicated by $\otimes$.

Attention Mechanism: The attention mechanism (AM) addresses the problem of information overload when computational resources are limited by allocating those resources to more critical tasks. A larger number of parameters allows the model to store more information and effectively represent features within neural networks, however, this could lead to information overload. Applying an AM helps the network prioritize relevant input data, filtering out non-essential information and focusing only on what is necessary for the task at hand. The attention value is determined through two processes: (1) defining attention distributions for each input dataset and (2) using these distributions to compute the weighted average of the input data. Equations (16) to (19) illustrate the computational concepts of the AM.

$$h_t = RNN(x_t, h_{t-1}) \tag{16}$$

$$c_i = \sum_{j=1}^{n} \alpha_{ij} h_j \tag{17}$$

$$\alpha_{ij} = \frac{exp\left(s(h_t, \overline{h_s})\right)}{\sum_{k=1}^{n} exp\left(s(h_t, \overline{h_s})\right)} \tag{18}$$

$$s\left(h_t, \overline{h_s}\right) = h_t, \overline{h_s} \tag{19}$$

The hidden state variable h_t is generated by merging the input x with the previous hidden state h_{t-1}. The c_i represents the weighted average of all hidden layers as well as the following hidden layer unit. The weight ratio of the hidden layer units is represented by α. Consequently, the target value is $\overline{h_s}$, where s represents the weight calculation method.

Proposed Model: Combining the AM into the LSTM network enables the model to achieve high prediction accuracy. Regardless of the input signal sequence length, conventional LSTM may transform it into a fixed length, which limits the model's ability to learn from extensive sequences. To enhance prediction accuracy, the AM fortifies connections between hidden layers and highlights significant information by assigning weight components. The network encoder utilizes the AM to retain intermediate outputs. The proposed model is trained to detect correlations between input and output sequences by selectively learning meaningful representations from the input sequence. Figure 5 illustrates how the AM computes the score for each variable by utilizing intermediate variables obtained from the hidden layers of the LSTM. The weight assigned to each variable reflects its relative importance. Ultimately, the vector layer captures important features from the input, integrates it, and assigns additional weight.

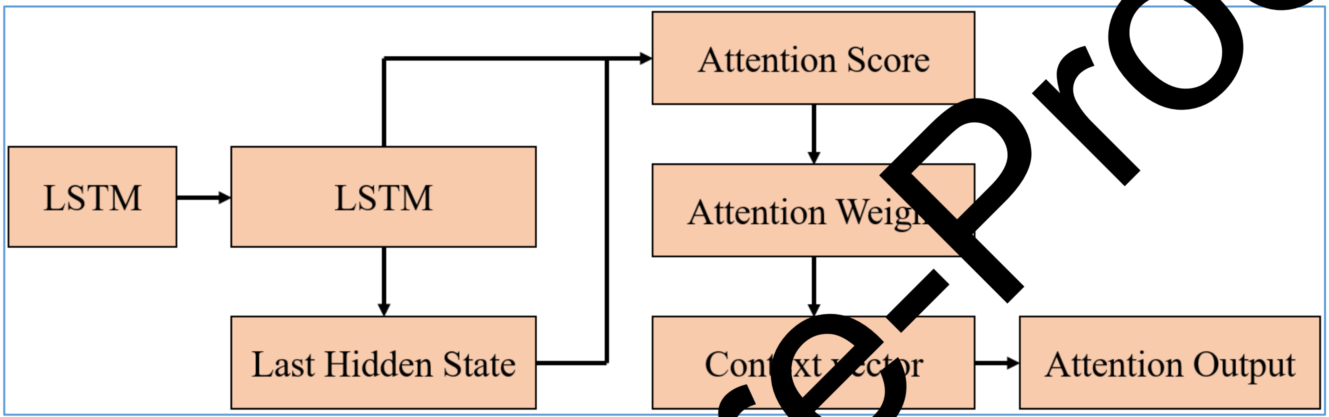


Fig. 2. LSTM-attention architecture.

After integrating LSTM with the AM, Equations (19)–(23) present the conditions for updating each parameter.

$$\begin{pmatrix} f^n_{(t)} \\ i^n_{(t)} \\ o^n_{(t)} \\ g^n_{(t)} \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ tanh \end{pmatrix} Z^n_{L+M} \begin{pmatrix} Ed^n_{t-1} \\ h^n_{t-1} \\ v_t \end{pmatrix} \tag{19}$$

$$v_t = \sum_i^N \beta_i^t x_i \tag{20}$$

$$s_i^t = W_s \, tanh\left(W_h h^n_{(t-1)} + W_x x_i + b_s\right) \tag{21}$$

$$\beta_i^t = \frac{exp(s_k^t)}{\sum_{k=1}^n exp(s_k^t)} \tag{22}$$

$$\sum_i^N \beta_i^t = 1 \tag{23}$$

where Z^n_{L+M} represents the n-th layer parameters, σ denotes the sigmoid function. The embedding matrix is indicated as Ed^n_{t-1}, with L referring to the LSTM dimension. The context vector v_t represents the relevant input vector at the time t, with M denoting the dimension of v_t. The relevance score s_i^t determines the attention weights β_i^t. The networks for the attributes $X = \{x_1, x_2, \ldots, x_N\}$ can obtain attention weights after acquiring their relevant scores. The weight matrices applied in the proposed network at the time t' are detailed in Equations (24-26).

$$w'_{t',ih} = w_{t'}, ih\beta_{t'} \tag{24}$$

$$w'_{t',oh} = w_{t'}, oh\beta_{t'} \tag{25}$$

$$w'_{t',fh} = w_{t'}, fh\beta_{t'} \tag{26}$$

Enhancing the weights allows the network to focus on capturing significant data. The LSTM-Attention mechanism improves prediction accuracy by identifying the most important temporal features from the limited ECG data. The algorithm 2 gives the working of LSTM-Attention model.

Algorithm 2: LSTM-Attention Model

Input: Sequence of input data $\boldsymbol{X}$. LSTM hidden states $\boldsymbol{h}$, Attention weights $\boldsymbol{\alpha}$

Output: Accuracy.

Step 1: LSTM Update

a. Forget gate: $\boldsymbol{f_t = \sigma(W_f . x_t + V_f . h_{t-1} + b_f)}$
b. Input gate: $\boldsymbol{i_t = \sigma(W_i . x_t + V_i . h_{t-1} + b_i)}$
c. Output gate: $\boldsymbol{o_t = \sigma(W_o . x_t + V_o . h_{t-1} + b_o)}$
d. Memory cell update: $\boldsymbol{c_t = f_t \odot c_{t-1} + i_t \odot tanh(W_c . x_t + V_c . h_{t-1} + b_c)}$
e. Hidden state update: $\boldsymbol{h_t = o_t \odot tanh(c_t)}$

Step 2: Attention Mechanism

a. Calculate hidden state at time $\boldsymbol{t}$: $\boldsymbol{h_t = RNN(x_t, h_{t-1})}$
b. Compute the attention score for each input: $\boldsymbol{s_i^t = W_s . tanh(W_h . h_{t-1} + W_x . x_t + b_s)}$
c. Calculate the attention weight $\boldsymbol{\beta_i^t = \frac{exp(s_i^t)}{\sum_{k=1}^{n} exp(s_k^t)}}$
d. Ensure that the attention weights sum to 1 $\boldsymbol{\sum_i^N \beta_i^t = 1}$

Step 3: Update the LSTM parameters using Attention

a. $\boldsymbol{w'_{t',ih} = w_{t'}, ih\beta_{t'}}$
b. $\boldsymbol{w'_{t',oh} = w_{t'}, oh\beta_{t'}}$
c. $\boldsymbol{w'_{t',fh} = w_{t'}, fh\beta_{t'}}$

Step 4: Evaluate the LSTM-Attention Model using accuracy
Step 5: Repeat Steps 2 and 3 if evaluation outcome not staisfied

C. *LSTM-Attention-based Meta-Learning*

The LSTM-Attention model captures the most important temporal features from ECG, which helps identify abnormalities in the ECG. The integration of LSTM-Attention in the meta-learning concept can handle a single model for a variety of task predictions using ECG data. This helps the model to learn from past data and adapt quickly and accurately to new tasks.

The process of LSTM-Attention-based meta-learning is detailed as follows: The ECG input is given to the LSTM model, which captures the temporal features that help understand the health condition of an individual. The AM refines the features and identifies the most important ones from the LSTM output by adjusting the weight metrics. LSTM gives importance to all features, but not all features contribute equally to the prediction task. LSTM-Attention is trained on each task, such as HRV prediction, arrhythmia prediction, and abnormalities. For each task, the model uses specific features from the ECG related to that task. Meta-learning is used during training to optimize the meta-parameters for quick adaptation to various tasks. If new ECG data with a new disease comes, the proposed framework can identify the new or unseen disease using the meta-parameters learned during training. Therefore, the framework predicts the new task using previous training and task knowledge with minimal data.

IV. Result and Discussion

The section details the experimental setup, data used in the research, and the evaluation of the proposed model in comparison with existing models.

A. *Experimental Set-up*

The experiment was conducted on the Google Colaboratory platform. The data are stored in Google Drive for easy access. The Graphics Processing Unit (GPU) available in Colab was used to run the code. The code is written in Python. The scikit-learn library is used for data processing and model evaluation. The model is imported from TensorFlow. The proposed method block diagram is given in Figure 3. The MIMIC-III data is collected and the required preprocessing is done. Then, the three different tasks are taken, and the LSTM-attention model is designed and tuned for each task individually using the meta-learning concept. If new data comes in, the proposed meta-learning-based LSTM-attention model can predict the disease accurately. The outcome of the model is evaluated using standard metrics.

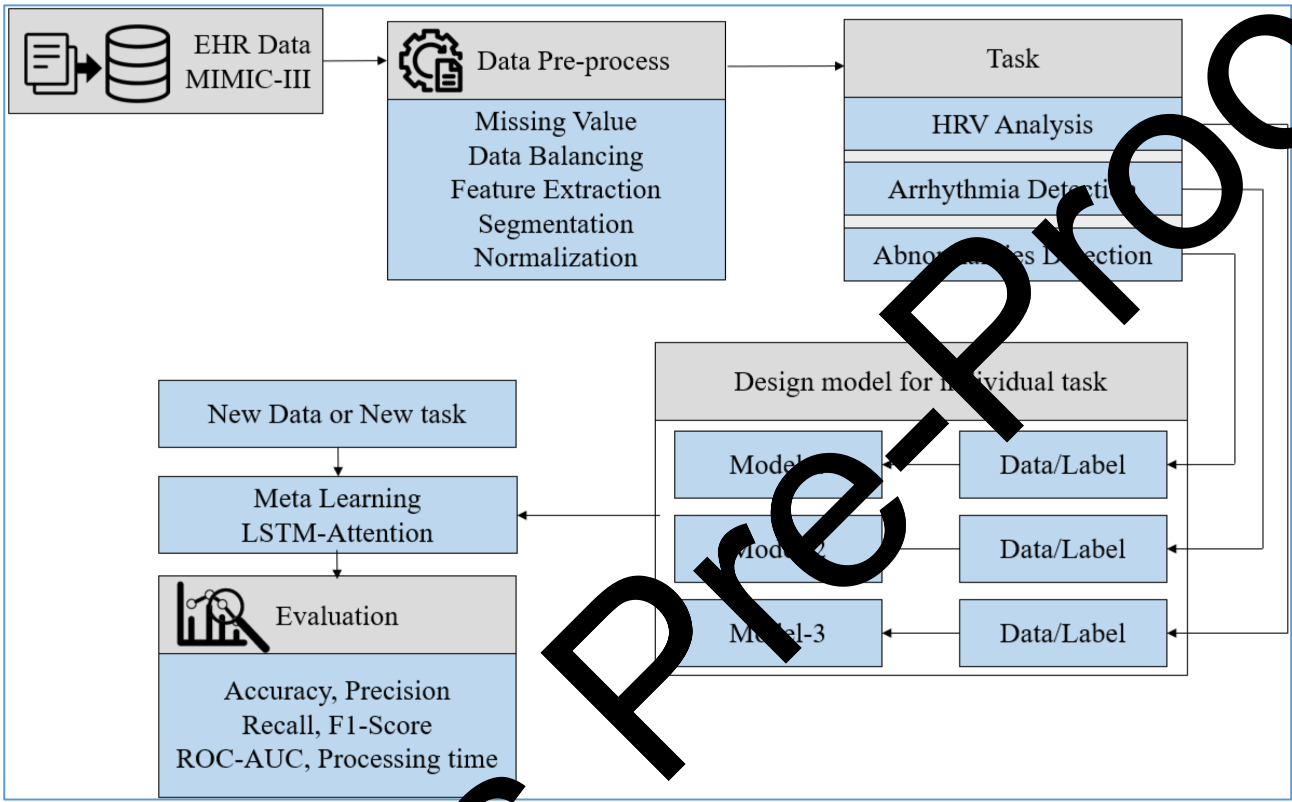


Fig. 3. Block diagram of the proposed method for disease prediction from EHR data.

B. Data Collection and Processing

The MIMIC III dataset [18], which includes medical information for intensive care units, was used in this pilot study. MIMIC-III is a database that contains the medical records of more than 60,000 individuals who were hospitalized in the intensive care units of Beth Israel Deaconess Medical Center between 2001 and 2012. This database holds extensive data. The study dataset primarily focuses on ECG data stored in EDF (European Data Format) files. The EDF format is widely used for storing time-series data from various biological signals. The main components of an EDF file include the metadata within the header section and the signal data in the main body of the file. In the context of the dataset, the EDF files likely contain multiple channels of ECG data collected over time. The ECG signal, representing the heart's electrical activity, is crucial for diagnosing and monitoring cardiac conditions. QRS points refer to annotations on the QRS complex in an ECG signal. Accurate recognition of the QRS complex is essential for tasks such as HRV analysis, arrhythmia detection, and other abnormalities.

The first pre-processing step is to identify the missing data and fill in the missing values using the imputation method. Next, the data using the Synthetic Minority Over-sampling Technique (SMOTE) method [19] to ensure the data is balanced, with each label in the dataset having equal samples. For feature extraction, the HRV distribution is calculated using standard deviation and rolling windows. Figure 4 shows the outcome of the HRV analysis and rolling window.

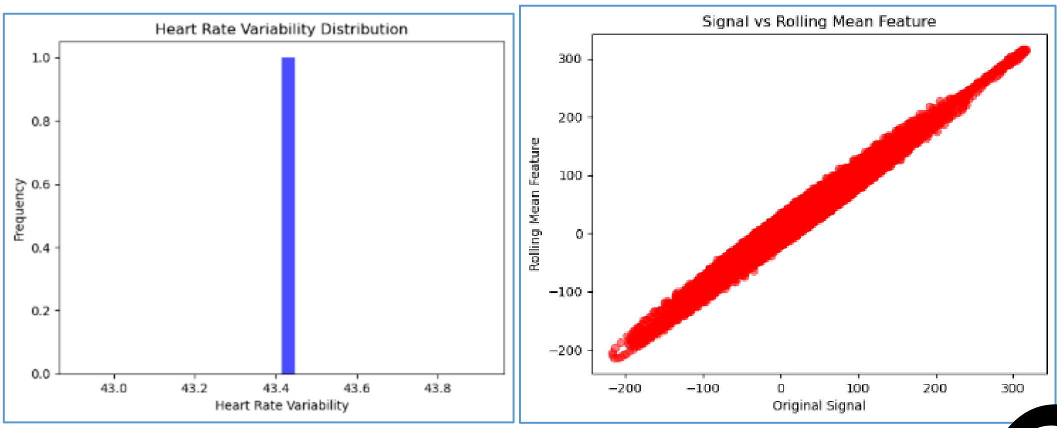


Fig. 4. a) HRV distribution b) Rolling Window.

Finally, segmentation and normalization are performed on the data. In segmentation, the continuous signal is divided into meaningful intervals. For normalization, the standardization technique is applied to convert the values of the data into a range from 0 to 1. This helps to reduce the memory usage and complexity of model training. Figure 5 provides the signal plot before and after normalization.

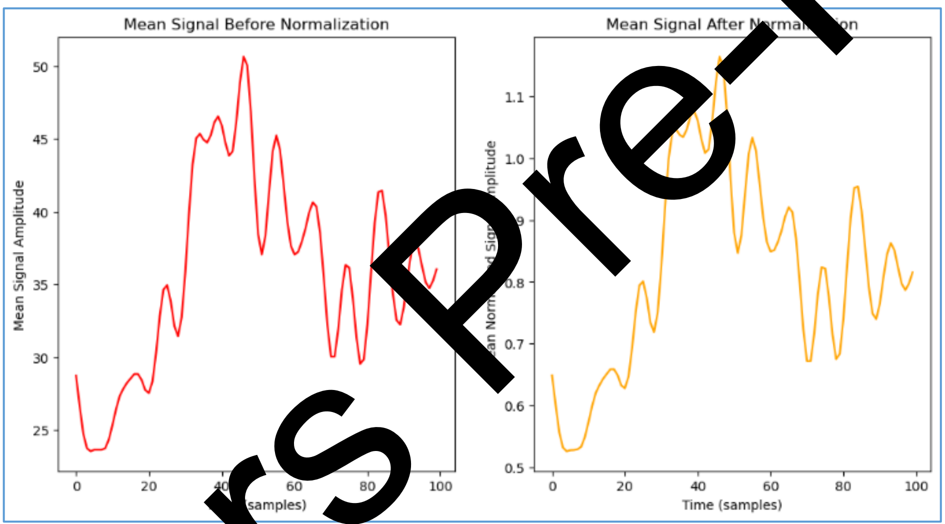


Fig. 5. Signal plot before and after normalization.

C. Result Comparison

After processing the data, it is provided to the meta-learning framework, which includes the LSTM-Attention model. To evaluate the effectiveness of the proposed LSTM-Attention model, it is compared with other models such as SVM, Random Forest (RF), and XGBoost. All the models are implemented within the meta-learning framework. The three tasks—HRV prediction, arrhythmia prediction, and abnormality prediction—are performed using the same dataset. The processed data is split into training, validation, and test sets in a 7:2:1 ratio. The LSTM-Attention model is trained on each task individually, and metrics such as accuracy and loss are used during the training and validation processes. The results of the LSTM-Attention model for each task are shown in Figures 6-8.

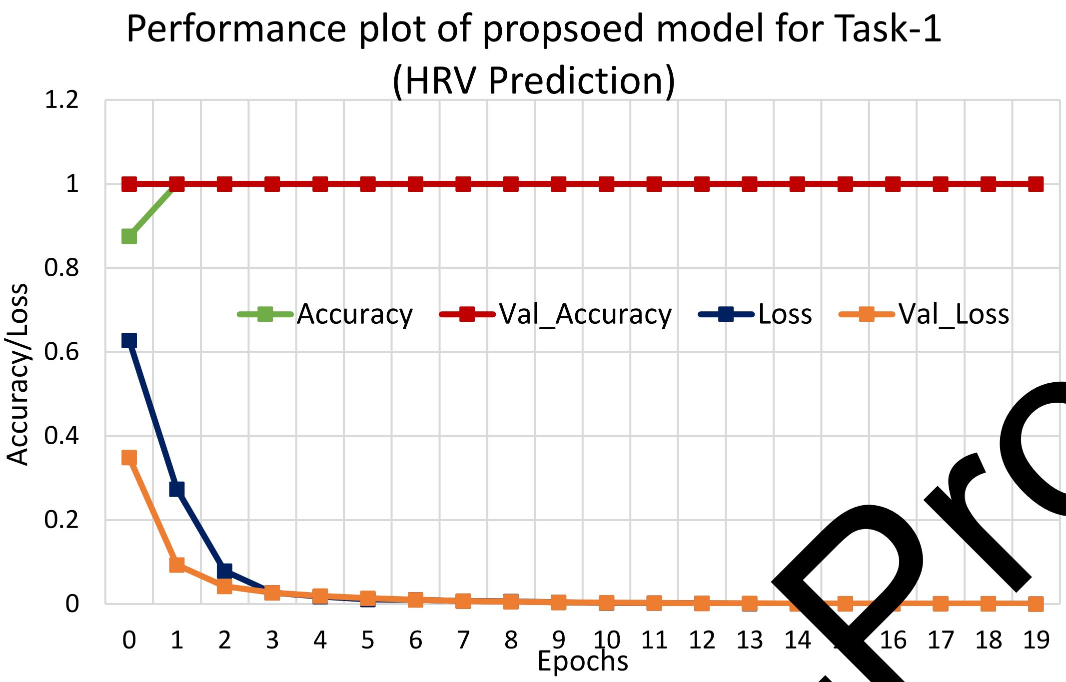


Fig. 6. Performance plot of propsoed model for HRV prediction.

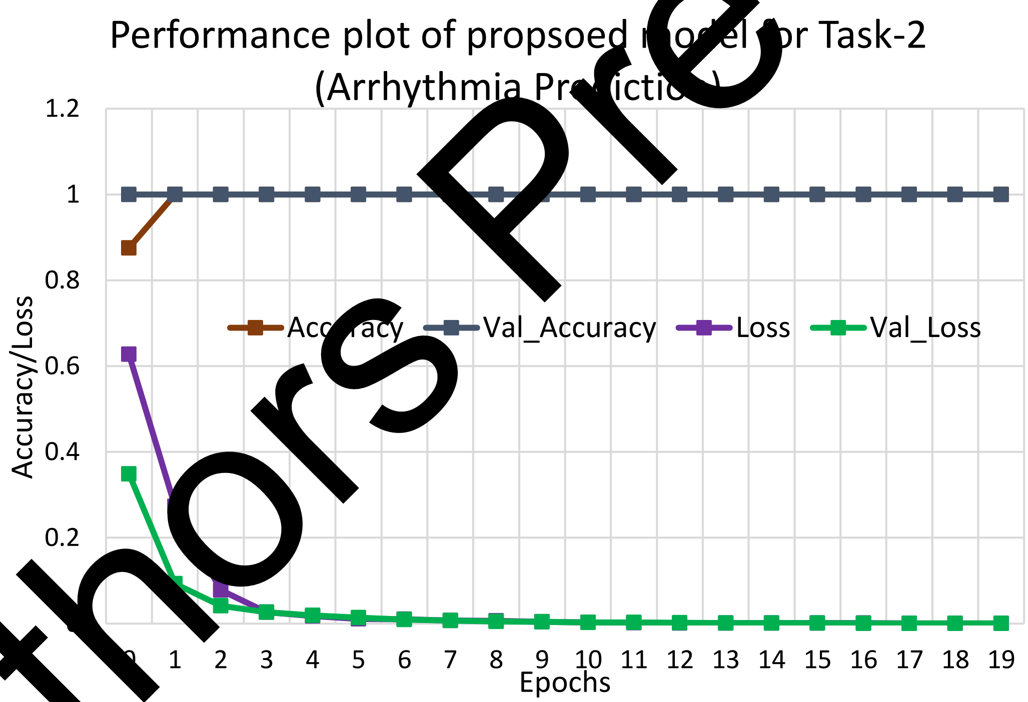


Fig. 7. Performance plot of propsoed model for Arrhythmia prediction.

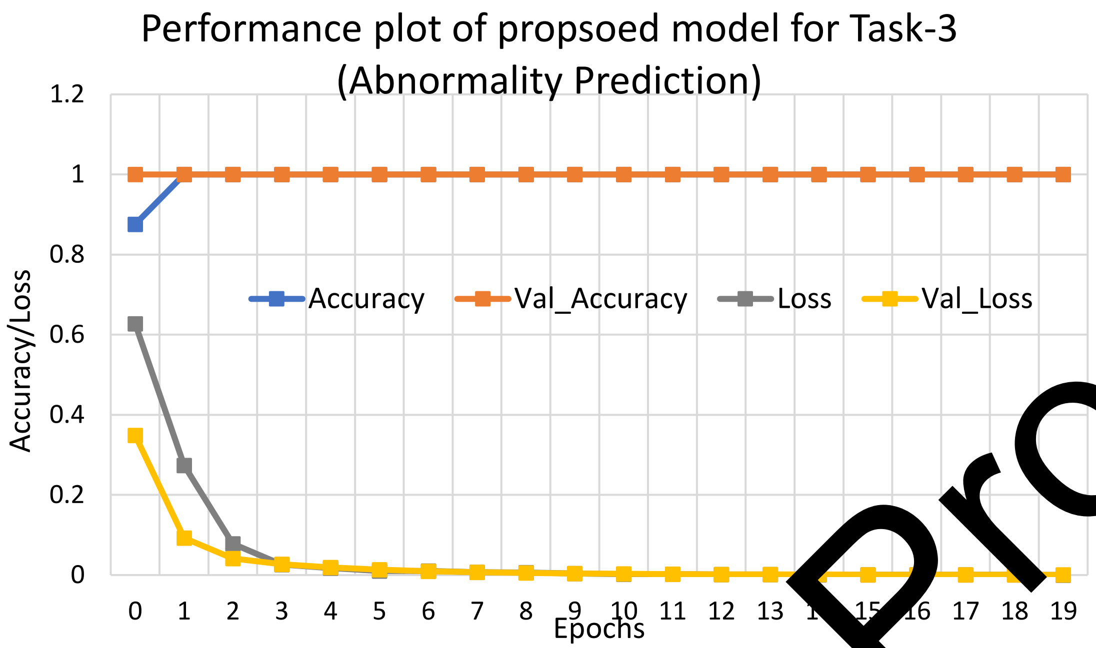


Fig. 8. Performance plot of propsoed model for Abnormality prediction.

After training and validation, the model is tested to evaluate its performance. The evaluation metrics such as accuracy, precision, recall, F1 score, and ROC-AUC are compared in addition to processing time. The highest metrics of accuracy, precision, recall, F1 score, and ROC-AUC are attained by the proposed model, with values of 0.92, 0.90, 0.91, 0.91, and 0.93, respectively. Next, the highest scores for accuracy (0.87), precision (0.83), recall (0.85), F1 score (0.84), and ROC-AUC (0.88) are obtained by RF. XGBoost shows better results, and finally, SVM performs the worst. The metric values for all models are shown in Table 1. The table also provides the formulas to calculate the metrics, where TP, TN, FP, FN represent the True Positive, True Negative, False Positive, and False Negative for the model's prediction.

Table 1. Performance comparison of proposed LSTM-Attention model with conventional models

Methods	Formula	SVM [20]	Random Forest [21]	XGBoost [22]	Proposed LSTM-Attention
Accuracy	$\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$	0.85	0.87	0.86	0.92
Precision	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$	0.81	0.83	0.82	0.90
Recall	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$	0.84	0.85	0.83	0.91
F1 Score	$\frac{2 * \mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}}$	0.82	0.84	0.83	0.91
ROC_AUC	$\int_0^1 \left(\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\right) d\left(\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}\right)$	0.86	0.88	0.87	0.93
Processing Time	-	120	115	110	95

Finally, processing time is important when deploying the model in real-time applications. The proposed model takes just 95 seconds to predict the disease from ECG data, while the other models—SVM, Random Forest, and XGBoost—take 120 seconds, 115 seconds, and 110 seconds, respectively. The quality metrics show the effectiveness of the proposed model in disease prediction, and the processing time indicates that the model can be deployed in real-time applications. Figures 9–14 show the comparison plots for the quality metrics and processing time attained by the models. The visual graphs help in a better understanding of the model evaluation.

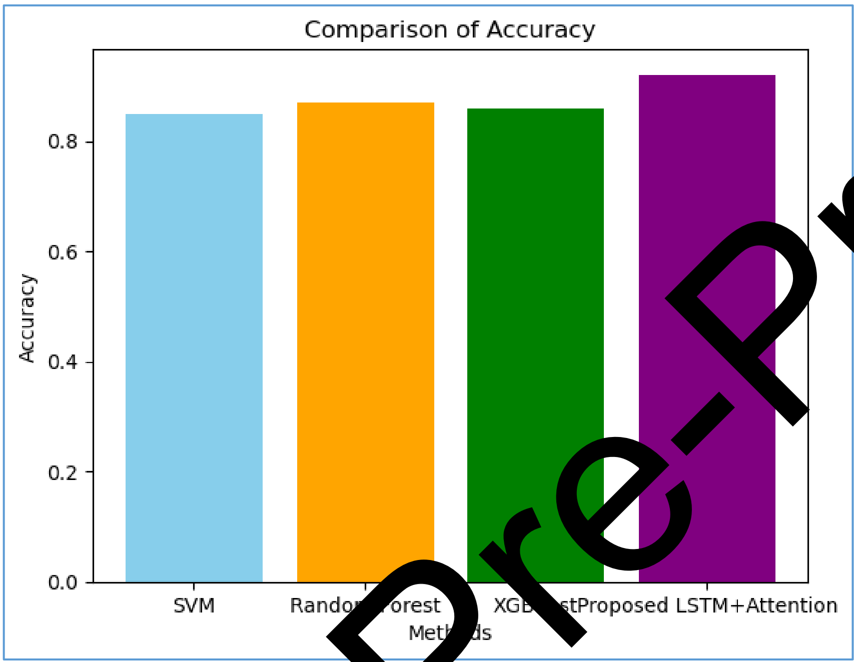


Fig. 9. Accuracy comparison of proposed LSTM-Attention-based Meta-learning model with conventional models.

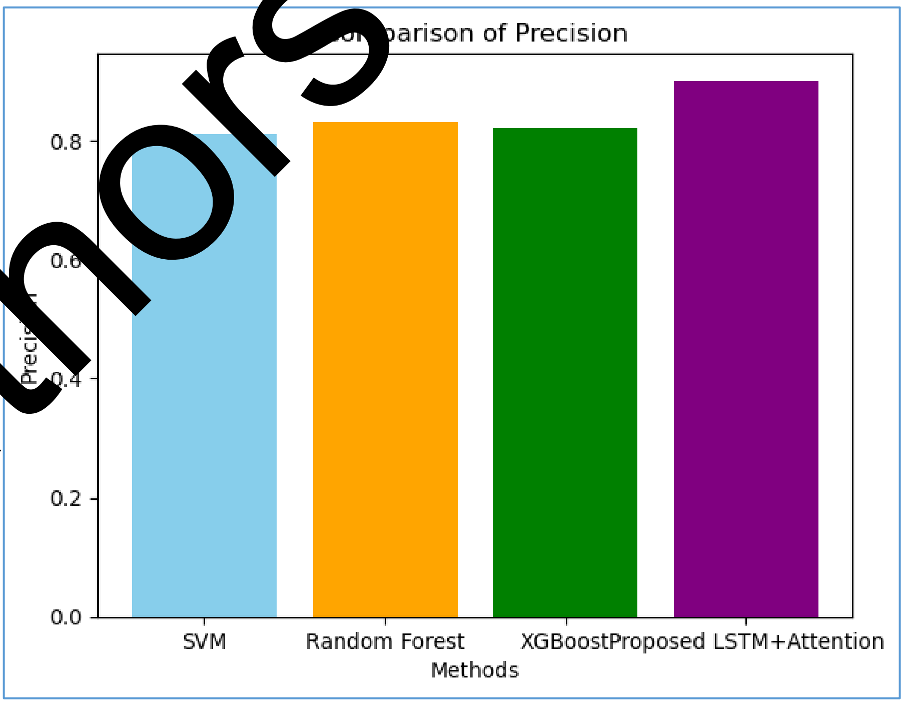


Fig. 10. Precision comparison of proposed LSTM-Attention-based Meta-learning model with conventional models.

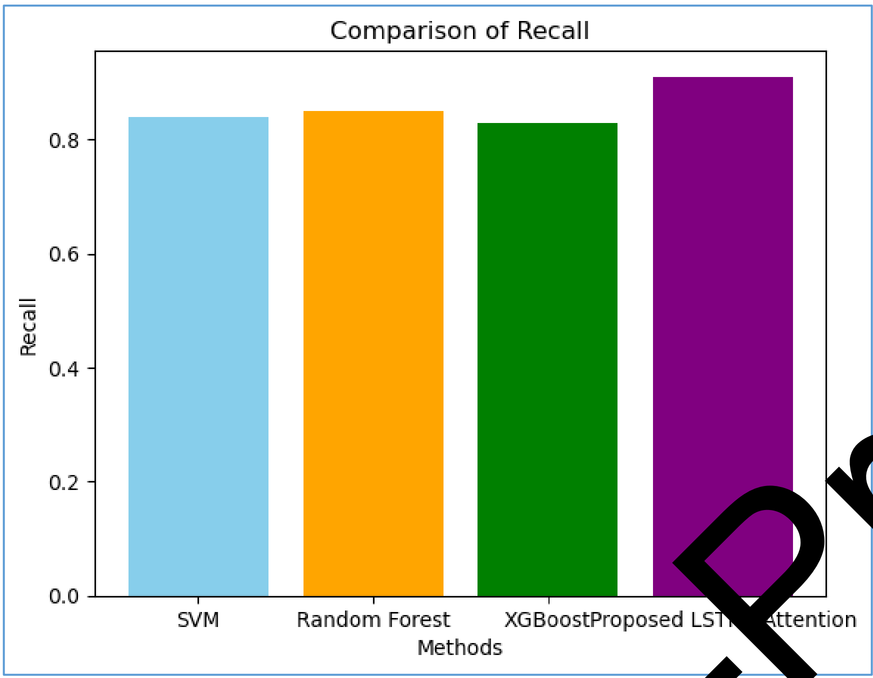


Fig. 11. Recall comparison of proposed LSTM-Attention-based Meta-learning model with conventional models.

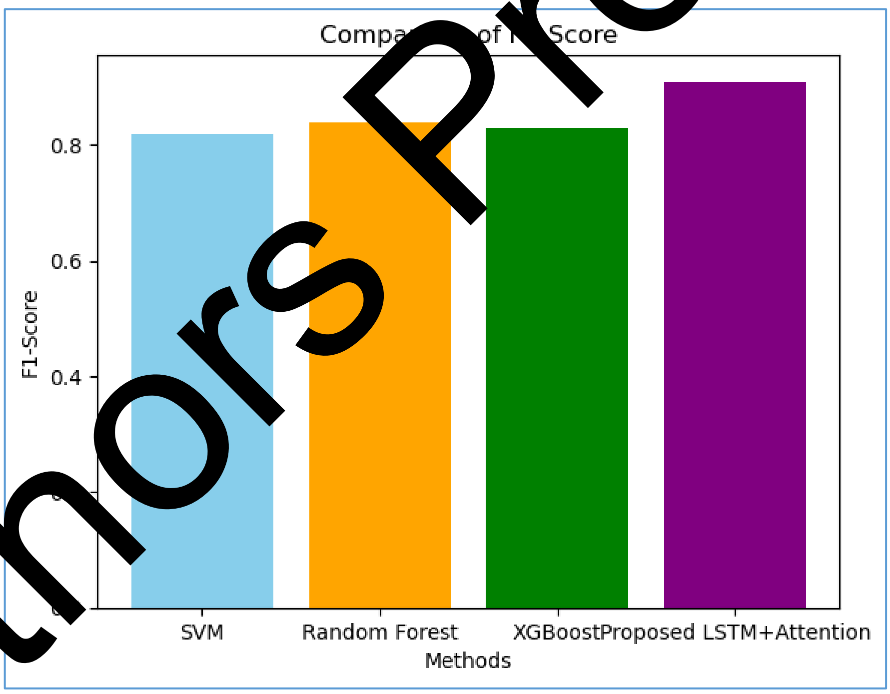


Fig. 12. F1-Score comparison of proposed LSTM-Attention-based Meta-learning model with conventional models.

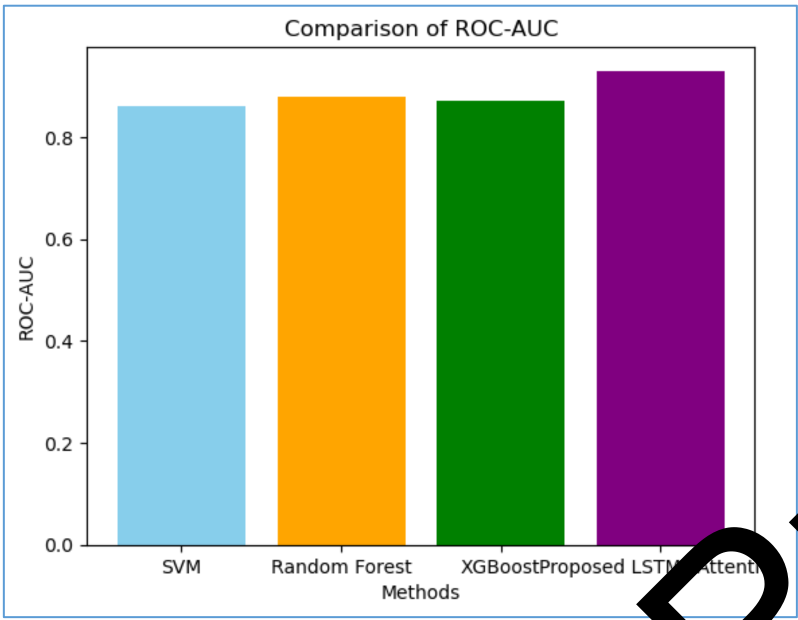


Fig. 13. ROC-AUC comparison of proposed LSTM-Attention-based Meta-learning model with conventional models.

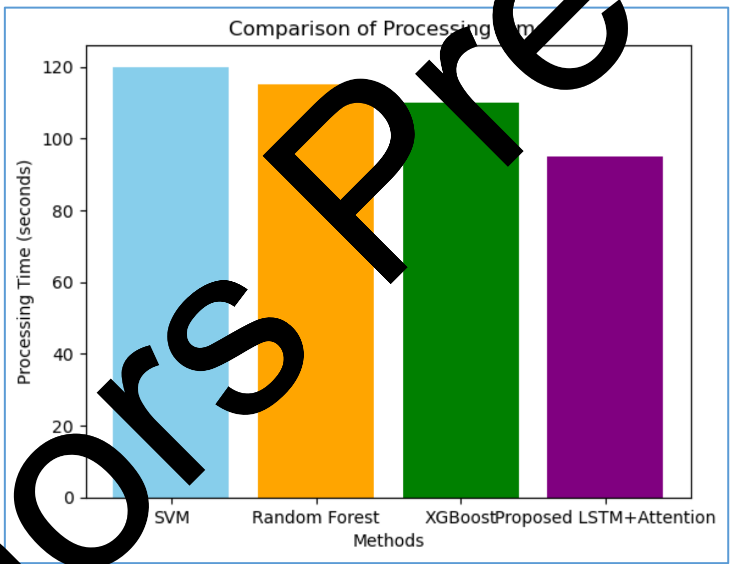


Fig. 14. Processing time comparison of proposed LSTM-Attention-based Meta-learning model with conventional models.

V. Conclusion

The research aims to propose an efficient model for predicting all types of diseases from ECG data. While many research models are available, they are often generalized and fail to account for unique patient data, particularly when dealing with rare diseases. To address this issue, the research proposes a novel LSTM-Attention-based meta-learning model for predicting diseases from ECG signals in EHR data. The ECG data in EHR is often limited, but by using the proposed model, it is capable of predicting three different tasks: HRV, arrhythmia, and abnormality prediction. The integration of AI models and meta-learning can solve many problems in the healthcare field by providing accurate predictions with limited data, unseen data, etc. The proposed model is compared with SVM, RF, and XGBoost, and it achieves the highest accuracy of 0.92, outperforming RF with an accuracy of 0.87. The proposed model improves prediction accuracy by 5%.

Additionally, compared to other models, the proposed model requires minimal time (95 seconds) to predict diseases from ECG data. The quality and processing time of the model ensure that it can be implemented in clinical applications. The proposed framework will help healthcare professionals predict and treat diseases as early as possible.

In the future, to enhance the accuracy of the proposed model, multi-modal data will be used. Other data, such as medical imaging, will be incorporated with ECG data to provide a comprehensive health profile. To deploy the proposed model in real-time, it will be further refined for integration into a wearable device for continuous health monitoring and immediate disease prediction.

Declaration

Ethics approval and consent to participate
Not applicable.

Clinical trial number
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Data Availability
The dataset generated and/or analysed during the current study is available on MIMIC III dataset [27], https://mimic.mit.edu/ .

Authors Contribution
C.G., S.M.B., S.T.A., S.K.M. participated in designing the methodology, concept and implemented the code and performed experiments, and wrote the manuscript. S.M. validated, supervised, reviewed and edited the manuscript.

Funding

The authors received no funding from their institutes.

References

[1] Vos, Janita FJ, Albert Boonstra, Arjen Kooistra, Marc Seelen, and Marjolein Van Offenbeek. "The influence of electronic health record use on collaboration among medical specialties." *BMC health services research* 20 (2020): 1-11.

[2] Poongodi, T., D. Sumathi, P. Suresh, and Balamurugan Balusamy. "Deep learning techniques for electronic health record (EHR) analysis." *Bio-inspired Neurocomputing* (2021): 73-103.

[3] Tayefi, Maryam, Phuong Ngo, Taridzo Chomutare, Hercules Dalianis, Elisa Salvi, Andrius Budrionis, and Fred Godtliebsen. "Challenges and opportunities beyond structured data in analysis of electronic health records." *Wiley Interdisciplinary Reviews: Computational Statistics* 13, no. 6 (2021): e1549.

[4] Cheng, Yu, Fei Wang, Ping Zhang, and Jianying Hu. "Risk prediction with electronic health records: A deep learning approach." In *Proceedings of the 2016 SIAM international conference on data mining*, pp. 432-440. Society for Industrial and Applied Mathematics, 2016.

[5] Jiang, Song, Yuan Gu, and Ela Kumar. "Stroke risk prediction using artificial intelligence techniques through electronic health records." *Artificial Intelligence Evolution* (2023): 88-98.

[6] HL, Gururaj, Vinayakumar Ravi, Meshari Almeshari, and Yasser Alzamil. "Electronic health record (EHR) System development for study on ehr data-based early prediction of diabetes using machine learning algorithms." *The Open Bioinformatics Journal* 16, no. 1 (2023).

[7] Tsien, Edward, Dezhi Wu, and Ana Lopez-De Fede. "Developing Multi-Task Learning Methods to Aid in Electronic Healthcare Prediction." In *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)*, pp. 486-487. IEEE, 2023.

[8] Suo, Qiuling, Fenglong Ma, Ye Yuan, Mengdi Huai, Weida Zhong, Aidong Zhang, and Jing Gao. "Personalized disease prediction using a CNN-based similarity learning method." In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 811-816. IEEE, 2017.

[9] Liu, Kang, Xiangzhou Zhang, Weiqi Chen, S. L. Alan, John A. Kellum, Michael E. Matheny, Steven Q. Simpson, Yong Hu, and Mei Liu. "Development and validation of a personalized model with transfer learning for acute kidney injury risk estimation using electronic health records." *JAMA Network Open* 5, no. 7 (2022): e2219776-e2219776.

[10] Tian, Jingxiao, Ao Xiang, Yuan Feng, Qin Yang, and Houze Liu. "Enhancing Disease Prediction with a Hybrid CNN-LSTM Framework in EHRs." *Journal of Theory and Practice of Engineering Science* 4, no. 02 (2024): 8-14.

[11] Dong, Xinyu, Rachel Wong, Weimin Lyu, Kayley Abell-Hart, Jianyuan Deng, Yinan Liu, Janos G. Hajagos, Richard N. Rosenthal, Chao Chen, and Fusheng Wang. "An integrated LSTM-HeteroRGNN model for interpretable opioid overdose risk prediction." *Artificial intelligence in medicine* 135 (2023): 102439.

[12] Poongodi, T., D. Sumathi, P. Suresh, and Balamurugan Balusamy. "Deep learning techniques for electronic health record (EHR) analysis." *Bio-inspired Neurocomputing* (2021): 73-103.

[13] Kiser, Amber C., Karen Eilbeck, and Brian T. Bucher. "Developing an LSTM model to identify surgical site infections using electronic healthcare records." *AMIA Summits on Translational Science Proceedings* 2023 (2023): 330.

[14] Liu, Siru, Joseph J. Schlesinger, Allison B. McCoy, Thomas J. Reese, Bryan Steitz, Elise Russo, Brian Koh, and Adam Wright. "New onset delirium prediction using machine learning and long short-term memory (LSTM) in electronic health record." *Journal of the American Medical Informatics Association* 30, no. 1 (2023): 120-131.

[15] Jabir, K., and A. Thirumurugan Raja. "Prediction of Lung Cancer from Electronic Health Records Using CNN Supported NLP." In *Computational Intelligence for Clinical Diagnosis*, pp. 549-560. Cham: Springer International Publishing, 2023.

[16] Zou, Lan. *Meta-learning: Theory, algorithms and applications*. Elsevier, 2022.

[17] Sutskever, I. "Sequence to Sequence Learning with Neural Networks." *arXiv preprint arXiv:1409.3215* (2014).

[18] MIMIC III dataset. Available Online: https://mimic.mit.edu/ (access date: 5 April, 2024).

[19] Adi Pratama, Firza Refo, and Siskarossa Ika Oktora. "Synthetic Minority Over-sampling Technique (SMOTE) for handling imbalanced data in poverty classification." *Statistical Journal of the IAOS* 39, no. 1 (2023): 233-239.

[20] Cortes, Corinna. "Support-Vector Networks." *Machine Learning* (1995).

[21] Breiman, Leo. "Random forests." *Machine learning* 45 (2001): 5-32.

[22] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794. 2016.

Appendix 1. Nomenclature

Symbol	Description
$\boldsymbol{x}$	Observation/input data
$\boldsymbol{y}$	Corresponding label/output
$\boldsymbol{D}$	Training data set
$\boldsymbol{\phi}$	Model parameters
$\boldsymbol{p(\phi)}$	Prior probability of the model parameters
$\boldsymbol{D_{meta-train}}$	Meta-training data set
$\boldsymbol{\theta}$	Meta-parameters learned from the meta-training data
$\boldsymbol{\theta^*}$	Point estimate of the meta-parameters
$\boldsymbol{f_t}$	Forget gate in LSTM
$\boldsymbol{i_t}$	Input gate in LSTM
$\boldsymbol{o_t}$	Output gate in LSTM
$\boldsymbol{c_t}$	Cell state in LSTM
$\boldsymbol{h_t}$	Hidden state in LSTM
$\boldsymbol{\sigma(x)}$	Sigmoid activation function
$\boldsymbol{h_t}$	Hidden state vector generated by RNN at time step
$\boldsymbol{\alpha_{i,j}}$	Attention weight for the i-th element and j-th time step
$\boldsymbol{\alpha_{i,j}}$	Normalized attention weight
$\boldsymbol{s(h_t, h_s)}$	Similarity function between hidden states
$\boldsymbol{v_t}$	Context vector at time t in AM
$\boldsymbol{\beta_i^t}$	Attention weight at time t for the i-th input
$\boldsymbol{w_{(t'),ih}}$	Weight matrix for input-hidden layer connections at time t'
$\boldsymbol{w_{(t'),oh}}$	Weight matrix for output-hidden layer connections at time t'
$\boldsymbol{Z^n_{(L+M)}}$	n-th layer parameters of the LSTM-Attention network
$\boldsymbol{W_s}$	Weight matrix for the similarity function in AM
$\boldsymbol{W_h}$	Weight matrix for hidden state in AM
$\boldsymbol{W_x}$	Weight matrix for input in AM
$\boldsymbol{b_s}$	Bias term for the similarity function in AM