Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 7 (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Optimizing Feature Selection for Deep Learning Models in Heart Disease Prediction Using ECG Data

Saroj Kumari[1*], Raghav Mehra[2]

[1*]Research Scholar, Department of Computer Engineering & Applications, Mangalayatan University, Aligarh, UP, India.
Email ID: saroj.cse10@gmail.com
[2]Professor, Department of Computer Engineering & Applications, Mangalayatan University, Aligarh, UP, India.
Email ID: raghav.mehrain@gmail.com

.Cite this paper as: Saroj Kumari, Raghav Mehra, (2025). Optimizing Feature Selection for Deep Learning Models in Heart Disease Prediction Using ECG Data. *Journal of Neonatal Surgery*,14 (7), 73-84

ABSTRACT

Heart disease popularly known as cardiovascular diseases (CVDs), continue to rank among the world's top causes of death, demanding the great need of designing accurate and effective diagnostic techniques. Early detection and accurate assessment of heart conditions can significantly enhance patient outcomes and lower the cost of healthcare. In this paper we have investigated the impact of machine learning (ML) models that use ECG data to predict heart disorder/diseases are affected by advanced feature selection techniques. Utilizing a dataset of 986 patients, the study focused on important features extracted using methods namely- Mutual Information (MI), Recursive Feature Elimination (RFE), L1 Regularisation, and Principal Component Analysis (PCA). The deep learning models involving are CNN, LSTM, MLP and ViT were investigated. With an accuracy of 99.30%, CNN with MI produced results that were competitive among the evaluated configurations, while LSTM with MI showed the best accuracy of 99.07%. With the ten feature selection approach-Top, CNN becomes the best model there also for the task at hand with the accuracy of 99.07%. The MLP and ViT models also performed well, achieving high precision (~97.74%) and accuracy of 97.67%. This comparative analysis emphasizes the significance of feature selection in reducing dimensionality, increasing computing effectiveness, and improving the model's performance. These finding highlight the potential of integrating advanced feature selection techniques with machine learning algorithms to detect cardiac disorders early. In order to increase clinical usability and ensure reliable performance across a variety of datasets, future research may explore hybrid models and real-time applications.

Keywords*: Cardiovascular disease, Machine learning, Electrocardiogram (ECG), Heart disease detection, Feature extraction.*

1. INTRODUCTION

A major and leading cause of death that took lives of millions of people is heart disease. The World Health Organisation (WHO) gauged, heart disease causes one-third of all fatalities worldwide, or over 18 million deaths [1]. This count reports for almost 30% of global deaths. Approximately 4% of yearly healthcare expenditures are spent on treating heart disease, and over 55% of deaths of cardiac patients occur within the first three years [2]. If appropriate action has not been taken on time, then it is predicted that the deaths from cardiac illness would rise to 22 million by 2030 [3]. With increasing prevalence of cardiac illness, timely and accurate detection is essential for effective treatment and better management of healthcare resources. Early diagnosis helps prevent complications and reduces the burden on medical systems. Several factors contribute to heart disease risk, including lifestyle habits, age, gender, and smoking. Recognizing these risk factors allows for early interventions which cut off the overall expenses of healthcare and enhance patient condition. These conditions are brought on by aberrant heart function and can be impacted by factors such as body weight, poor diet, alcohol consumption, high blood fat, blood sugar, obesity, and family history. Understanding the warning signs and behaviours of heart problems is essential. To detect cardiac diseases, a number of tests are necessary, counting with blood pressure, blood sugar, fat, auscultation, and electrocardiograms. Setting these tests as a top priority is essential since they frequently require long period of time to finish, and the patient must begin following prescription immediately. Understanding the more healthy behaviour that lead to heart diseases is essential [4]. However, due of the many risk factors, this disorder is difficult to diagnose [5]. Better patient outcomes are greatly aided by early identification and thus early detection of cardiac disorders is essential for averting serious consequences and lowering death rates.

Thanks to the impressive recent advancements in machine learning (ML), particularly in the medical field, clinicians now have a better understanding of how to gather and interpret data. Artificial intelligence and ML have number of applications in different area, such as smart support, workplace safety, and health monitoring. In order to identify anomalous conditions (such as heat stress, falls, or exposure to dangerous gases), ML models analyse data using trends in input data to forecast [6]. Incredible machine learning analysis, namely in the area of electrocardiogram (ECG) diagnosis, has enabled rapid interpretation employing ECG characteristics as the ideal foundation for this process, because it can non-invasively track heart's electrical activity and make a record of it, electrocardiograms (ECGs) are among the most popular and efficient methods for identifying cardiac disorders. Many heart abnormalities like arrhythmias, myocardial infractions and early indicators of heart disease and heat conditions can be easily recognized and identified by clinicians, with examining ECG data [7]. However, because heart diseases are of many different types, sometimes it becomes very challenging to diagnose heart disease even for a highly qualified cardiologist, based solely on an ECG signal. Furthermore, due to the resemblance of signals, it might be very difficult to accurately distinguish between the symptoms of cardiac illness on ECG readings. Also, ECG signal recording may have variances and differences for the same medical condition, depending on age, race, and overall physical condition of the patient's [8, 9].

Machine learning in ECG analysis has shown a lot of promise in recent year to increase the accuracy and efficacy of cardiac disease identification. This is because the intricate patterns in ECG data would be undetectable to the naked eye but machine learning algorithms' has the capability to identify such intricate patterns in ECG data. However, the calibre and applicability of the features that are taken from ECG signals have a significant impact on how well these models work. Choosing the right features is a crucial part of the machine learning process because high dimensional, noisy, or irrelevant features can impair model performance. Since ML can identify these patterns, it has been used to a variety of heart conditions, such as cardiomyopathy, atrial fibrillation, and left ventricular systolic dysfunction. Using a huge patient's data, ML techniques have lately shown encouraging leads to the forecasting of heart conditions [10]. Deep Learning (DL) techniques, namely Convolutional Neural Networks (CNN), Long Short Term Memory (LSTM), Vision Transformer (ViT) and Multi Layer Perceptron (MLP) have made it possible to directly learn complex patterns from raw data. DL models especially LSTM and CNN are very powerful in identifying irregularities over time which are useful for modelling sequential data, such as time-series logs [11]. These models can be used to identify and analysing patterns specially CNN model uses several layers to process the image namely Convolutional Layer that is used to extracts an object's edges, forms, and textures, Pooling Layer which preserves key characteristics while reducing dimensionality and the Fully connected Layer which classify data through retrieved features in data [12].This makes them perfect for jobs like image based analysis and diagnosis.

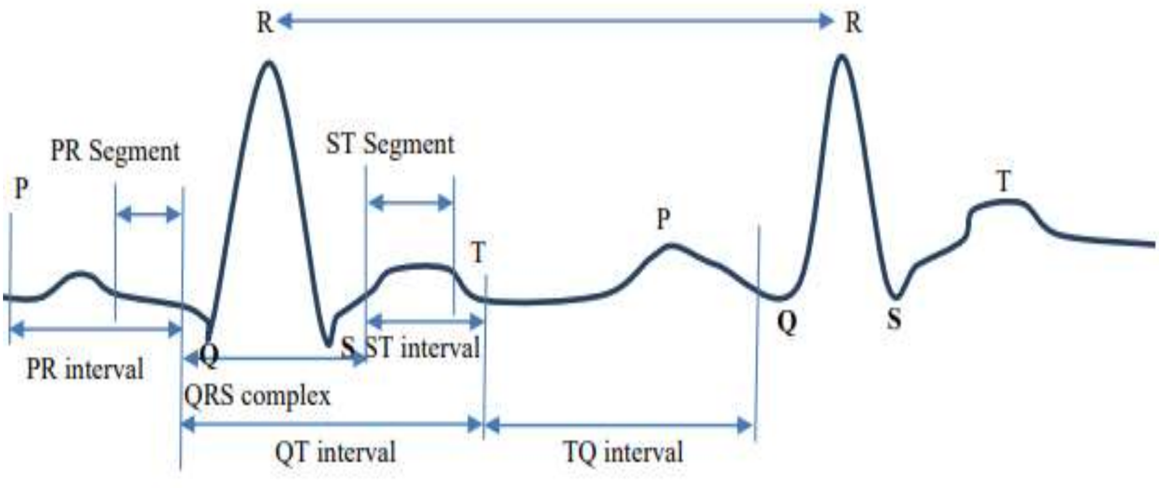

Figure1: General ECG signal and fiducia [13]

The signal produced by an ECG may contains a range of different waveform or heartbeats, as seen in Figure 1, therefore feature extraction and its proper selection from the input data is crucial. Following processing, the ECG output signal is containing many waves that collectively indicate cardiac problems in humans. The Q, R, S, P, T, and so forth waves are examples of these waves. The waves that make up the QRS complex- Q, R, and S, are focused on ventricular depolarisation. The T wave is in charge of ventricular repolarisation, while the P wave controls atrial depolarisation. Following pre-processing, feature extraction and characterisation are necessary for ECG-based heart disease prediction. The ventricles' depolarisation is controlled by the QRS complex. Atrials are depolarised by P waves, whereas ventricles are repolarised by T waves. Frequency space, also known as time frequency area, can be applied to show the different components of time and the frequency in the ECG simultaneously when QRS-complex power spectra are compare for arrhythmia wave and normal waveforms [14]. The quality of the input features significantly effect to the effectiveness of machine learning models in ECG-based heart disease identification. These features, such as morphological traits (namely- P-wave and QRS complex), time-domain traits (namely- R-R interval), and frequency-domain traits (like the power spectral density) represent important information that was taken from the ECG signals [15]. The huge dimensionality of raw ECG data and the existence of redundant, noisy, or irrelevant properties might impair machine learning model performance, and features extraction is more challenging than classification because QRS complexes are hard to comprehend [16, 17]. This challenge emphasises the importance of feature selection is necessary and important phase in the functioning of machine learning that comprises identifying the most relevant and instructional elements from the data. ML requires hand-crafted features, or traits that are produced via trial and error, to attain the maximum level of classification accuracy [18, 19]. In order to identify the important features and remove unnecessary and redundant features that affect the outcome of ML model prediction, the best perceivable features extraction and selection for ECG signal must be applied to carry out the accurate and pertinent classification. To enhance feature extraction, matching precision and computational efficiency, many studies suggests super symmetric classifiers which can enhance MRI, CT, and ultrasound images to enhance diagnosis [20]. Consequently, there are several reasons that feature selection is essential and important in ECG-based machine learning models. Primarily, it helps in improving the processing efficiency and reduce training time by reduce the dimensionality of the data, particularly for complex models such as MLP, CNN, LSTM, and ViT. These archetypes have revolutionised image processing by making automated feature extraction and pattern identification possible. This has greatly increased the accuracy of medical imaging, including the detection of abnormalities in CT, MRI, and X-ray images. These archetypes provide real-time analysis, automation, and increased accuracy, revolutionising image processing [21]. The second reason was that it lessens the likelihood of overfitting by eliminating superfluous or redundant characteristics that don't enhance the model's predictive power. The last reason was, by concentrating on the most essential and important aspects, feature selection might enhance the models' interpretability, which is crucial for clinical decision-making [22]. Some traits, for example, may be more resistant to modification in the patient's physiology or the conditions surrounding signal acquisition. Therefore, in ECG analysis, feature selection can also aid with problems related to signal variability and noise. The generalisability of machine learning models across various patient populations and datasets can be enhanced by choosing such features.

Larger datasets and the existence of multiple redundant and superfluous characteristics and features in the datasets represent significant weakness for the machine learning algorithms, despite the fact that machine learning is helpful for diagnosing a variety of ailments [23]. Additionally, sometimes often, only a few elements are necessary and relevant to the goal. When the remaining features are overlooked as redundant and superfluous, classification's capability and its accuracy degrade. Therefore, selecting a suitable and condensed subset of the basic features is essential to raise classification performance and triumph over the issues occurring due to high dimensionality. Goal of feature extraction and selection mechanism is to ascertain the importance of features. Feature selection significantly cuts down on processing time in addition to lowering the amount of inputs. Research on methodically assessing the effects of various feature selection strategies on ECG-based machine learning models is few; despite the crucial role that feature selection plays [24]. Prior research frequently concentrates on the use of certain methods without fully examining the ways in which feature selection affects model performance, computational effectiveness, and generalisability. This disparity emphasises the necessity of a thorough investigation that looks at the effects of feature selection strategies in relation to ECG data. Furthermore, effective feature selection techniques are becoming more and more crucial as ECG datasets get bigger and more complicated. The incorporation of sophisticated feature selection strategies, including filter, wrapper, and embedding methods, can greatly improve machine learning models' performance and allow them to identify minute patterns linked to heart disease in its early stages[25].

Despite their usefulness, a major shortcoming in the existing literature currently available on prediction of heart disease is the absence of standardized suggestions for choosing the features from input for prediction model creation, which is crucial for accuracy [26]. Prior study selected features primarily on the fly, without taking into account the most recent findings in medical research. While significant progress has been made in developing heart disease prediction models, there is still room for improvement. Various feature selection methods have been explored, yet refining these techniques further can enhance predictive accuracy. Resent research is needed to identify the most effective approaches for selecting relevant features, ensuring that prediction models deliver reliable and precise results. Prior studies on heart disease prediction have mostly concentrated on either optimising algorithms utilizing various machine learning approaches or trying to optimise techniques by utilizing variety of feature extraction and selection strategies. On measuring how various feature selection methods affect model’s performance however, has received less attention. How feature selection affects machine learning models' performance in ECG-based heart disease detection is the primary area this study aims to answer. In particular this study will try to find the solution for:

- Which feature selection technique work best to boost the efficiency and accuracy of ML model namely CNN, LSTM, MLP and ViT for ECG dataset to predict heart disease?
- What effects does feature selection have on the stability and interpretability of ML model utilised in ECG analysis?
- Does the identification of cardiac disease involve a trade-off between feature reduction and model performance?

  The answers of above findings is important from a technological and clinical standpoint in the following ways:

Enhancing Diagnostic Accuracy: Machine learning algorithms can more effectively discover tiny patterns in ECG data that point to early indicators of heart disease by identifying the most pertinent aspects.

Lowering the Cost of Computation: Because feature selection reduces the dataset's dimensionality, the models can be used in settings with limited resources, like wearable technology or mobile health apps.

Improving Clinical Applicability: Clinicians often used models that are more interpretable. Feature selection help in giving models which can be easily understandable and more reliable for decision making to clinical situations by focusing on only required and essential ECG values.

Resolving Data Challenges: Feature selection techniques are the prime concern of this research, which focuses on how different feature selection techniques handle the major challenges of ECG datasets, which includes factors namely- inconsistencies, redundant data and signal noise. Reliability and applicability of predictive models can be improved by focusing on important and critical features which produces accurate and interpretable insight in detection of heart diseases.

2. RELATED LITERATURE REVIEW

In the existing body of knowledge, numbers of studies have communicated the role and importance of using feature selection in ML models for the prediction and identification of diseases of heart based on ECG dataset. Instead of giving promising and good result in the identification of heart diseases, it is difficult for ML models to handle and manage high dimensional data such as ECG. Lowering the computational cost, finding the most relevant features and improving the model correctness, all depends on features selection. The models’ efficiency and interpretability is improved by using suitable feature selection techniques as these techniques concentrating on important ECG characteristics, which increases their usefulness in clinical settings.

To improve the ECG based heart disease detection many studies have explored different feature selection techniques. Disparate mechanism for selecting features including embedding, wrapper and filter techniques and their outcome on model accuracy was investigated by **Acharya et al.** [27]. In their study they found that ANOVA, Pearson correlation, and mutual information improved the model's performance significantly. With an accuracy of 94.03%, a deep CNN is used to classify arrhythmias. The finding of the study demonstrated that feature selection is a crucial pre-processing step in the interpretation of ECG data that lowers the likelihood of overfitting and improves model reliability. **Yildirim et al.** [28] applied RFE as a wrapper technique in another study to identify the most relevant ECG features. Their study significantly improved the classification of ECG data into various arrhythmia classifications by using deep learning models and Support Vector Machines (SVM) for feature extraction and selection. Their result demonstrated that systematically eliminating less significant features and enhanced performance and reduced data noise. To identify myocardial infraction in ECG data, different feature selection techniques were applied and compared in the study of **Sadhukhan et al.** [29]. RFE achieved an accuracy of 94.8%, LASSO Regression, an embedded technique, perform with highest accuracy of 96.4%, and the mutual information-based feature selection technique recorded an accuracy of 92.5%. Their results demonstrate that model’s accuracy and robustness improved by applying an advanced feature selection technique across many datasets. **Peker et al.** [30] in their study clarified the significance of using wrapper-based feature selection strategies in enhancing ML model performance in respect of accuracy in another significant study. Their results demonstrated that the classification accuracy increased from 85-90% without feature selection to 92-95% with using RFE and also feature selection reduces overfitting and enhances generalisation. **Mincholé et al.** [31] further explored the ECG-based cardiac illness models in conjunction with the feature selection approach to assess its impact on models’ performance, highlighting the need of selecting noise-resistant features to ensure model stability across a range of datasets. Their results confirmed the clinical appropriateness of ML models in the areaof cardiology and raised accuracy above 93% with the application of feature selection techniques. **Zhao et al.** [32] examined the effects of domain-specific feature selection in addition to traditional machine learning methods. The accuracy is improved from 85.7% to 92.4% by focussing on specific feature i.e heart rate variability (HRV), and their result highlights the importance of selecting clinically relevant features. To improve the ECG based model performance feature selection is still required and essential even with deep learning's advances. **Kiranyaz et al.** [33], proposed a hybrid strategy in order to enhance diagnostic performance and lower computing costs that combines CNNs with feature selection techniques. Their model achieves an accuracy of 98.5% by incorporating pertinent variables prior to training. **Liang et al. [**34] supported this approach, showing that pre-selected features helped CNN models achieve 94.6% accuracy compared to 89.3% when using raw ECG data. The importance of feature selection in ECG-based ML models continues to be a focal point of research. **Choi et al.** [35] used Random Forests to rank important ECG features, improving myocardial infarction detection accuracy to 92%. Similarly, **Chen et al.** [36] demonstrated that combining RFE with mutual information enhanced a Deep Neural Network's accuracy to 94.8%, underscoring the effectiveness of hybrid feature selection techniques.

While significant progress has been made, several gaps remain in feature selection for ECG-based models. Few studies have systematically compared feature selection methods across different ECG datasets and machine learning architectures. Additionally, limited research exists on optimizing feature selection for real-time ECG analysis in wearable devices or bedside monitors. More work is also needed to increase the clarity and make more understandable ML archetype for clinical executive. Moreover, advanced models like transformer-based architectures require tailored feature selection strategies to improve their applicability in cardiac diagnostics. For enhancing the detection and prediction accuracy of heart diseases using ECG dataset, selection of appropriate model is crucial because the features set include both sequential and structured data. For ECG based heart disease detection, we have selected deep learning models namely-Vision Transformer (ViT), Long Short Term Memory (LSTM), Convolution Neural Network (CNN) and Multi Layer Perceptron (MLP). These architectures were selected because of their prowess in efficiently processing and interpreting ECG data. The ECG features such as heart rate variability, ST slope, and ECG measurements are easily extracted in this study by using CNNs which eliminates the need of manual selection and hence improves the classification accuracy. However, LSTM is ideal for the heart disease features having temporal relationship such as heart rate variability, stress levels, and sleep length vary over time. Model performance improved by LSTM by evaluating patterns in sequential data in identifying heart disease risk factors. Cholesterol, blood pressure, diabetes and medication history etc. representing the structured clinical data, are processed using MLP. A robust baseline model for categorising the risk of heart disease by integrating pertinent feature selection strategies is MLP. The self-attention mechanism of ViT makes it chosen to comprehend intricate correlations among a variety of health metrics. It efficiently processes categorical and numerical features, making it useful for analysing combined datasets that include lifestyle habits, family history, and chronic diseases. By integrating these architectures, we aim to develop a more accurate and efficient heart disease prediction model. This research focuses on optimizing model performance by leveraging structured and sequential data while addressing challenges related to feature selection and high-dimensional datasets.

3. RESEARCH METHODOLOGY

The proposed methodology utilised a primary dataset collected from patients from three different clinics. This dataset comprises information from 986 patients and includes 26 features pertaining to patient demographics, lifestyle, and ECG readings (as provided in Table 1). However, only top ten of these features were utilized by different feature selection methods (as discussed in section 3.1) for heart disease prediction, as the others were deemed less impactful. Also, top ten features consistently selected across all the feature selection methods namely- RFE, MI, L1 and PCA. This set of features is denoted by Top and this set of features is used to train the entire selected model namely CNN, LSTM, MLP and ViT and evaluated to check the performance of selected model. Prior to categorisation, the dataset was filtered and cleaned to remove any missing or superfluous variables, using the imputation method. Before selecting relevant features, the dataset was standardized to ensure consistency in scale. Each feature was adjusted in order to improve the stability of ML model with a mean of zero and a standard deviation of one. After then, we have split the dataset at random into two sets i.e 30% (251 records) was used for testing, while 70% (735 records) was used for training. The prediction model was developed and improved using the training data, guaranteeing that it could successfully generalise to new, untested scenarios.

Table 1: Feature Set Description

S.No.	Feature Set
1.	ST slope (slope of the peak exercise ST segment)
2.	Cholesterol (Serum cholesterol in mg/dl)
3.	Sex (0 = female; 1 = male)
4.	Oldpeak (ST depression induced by exercise relative to rest)
5.	Age (Patient's age)
6.	BPS Resting (Resting BP in mm Hg)
7.	Chest pain (Kind of chest pain)
8.	FBP (Fasting blood sugar > 120 mg/dl)
9.	Exercise angina (Exercise-induced angina)
10.	Heart rate Max (Maximum heart rate in beats/min)
11.	ECG Resting (Resting ECG outcomes)
12.	Diabetes (0 = No, 1 = Yes)
13.	Heart Rate Variability (ms)
14.	Body Mass Index (BMI) (BMI, calculated from height and weight)
15.	Smoking status (1 = smoker, 0 = non-smoker)
16.	Alcohol consumption (1 = regular drinker, 0 = non-drinker)
17.	Physical activity level (1 = sedentary, 2 = moderate, 3 = active)
18.	Family history of heart disease (1 = yes, 0 = no)
19.	Blood Oxygen Level (SpO2 %)
20.	Respiratory Rate (breaths/min)
21.	Dietary habits (1 = poor, 2 = average, 3 = good)
22.	Sleep duration (Average hours of sleep per day)
23.	Stress level (1 = low, 2 = medium, 3 = high)
24.	Chronic diseases (1 = has chronic diseases, 0 = no chronic diseases)
25.	Medication history (1 = on heart medication, 0 = not on medication)
26.	Target (Heart attack risk, 1 = yes, 0 = no)

3.1 Common Feature Selection Techniques Utilized

This section mainly focuses to the evaluation of impact of feature selections on the outcomes of machine learning archetype. This analysisi targeted to show how feature selection can improve interpretability, save training time, and enhance model accuracy while maintaining or even enhancing model's predictive outcome. The successful feature selection methods utilized in this study for ECG-based classification are:

a) **Mutual Information (MI):** It is a type of Statistical Method that quantifies the relationship between the target variable and the input features. Each feature provide some information and by measuring the amount of information that each feature gives about the outcome helps in choose the features that have the highest influence on heart disease prediction.

b) **Recursive Feature Elimination (RFE):** RFE is classified under Wrapper Methods. In this technique while training a model, it systematically eliminates the least significant features and repeatedly assessing feature relevance at every iteration. In diagnosis of heart disease from ECG related characteristic namely QRS complex duration and RR intervals, RFE is found very useful and effective.

c) **L1 Regularization (Lasso):** Lasso is categorized under the Embedded Method of feature extraction and selection technique. Since, ECG data are considered as high dimension data and Lasso technique is very effective when working with high-dimensional. It ensures that only the most important features are retained, lowering the impact and setting the coefficient to zero by penalising less significant features during training.

d) **Principal Component Analysis (PCA)** This technique is considered under the Reduction of Dimensionality of feature extraction and selection method. In PCA, dimensionalities are reduced while maintaining the most significant important information.

The above techniques are applied in this study to make sure that only the most relevant features are employed, so that model's performance, efficiency and diagnostic reliability of predicting heart disease must be improved.

3.2 Deep Learning Models for ECG Data Analysis

In this section, we provide an overview of the deep learning models that were implemented in accordance with their significance and applicability in managing time-series and tabular data, especially for ECG-based data for predicting heart disease.

a) **Convolutional Neural Networks (CNNs):** The spatial patterns are efficiently detected by CNN using its convolutional layers with filters and kernels. CNN is widely used in image processing; even also this model's convolutions have been adjusted for time-series data by using them along the temporal dimension. The best part of CNN is that it quickly recognized local correlations in ECG data, such as variations in heart rate and ST slope [37].

b) **Long Short-Term Memory (LSTM):** LSTM is considered as a kind of recurrent neural network that utilises memory cells and gating mechanism to manage sequential dependencies in input data, this aids in identifying how certain ECG features namely- heart rate maximum, oldpeak, and ST slope changes, change over time. Heart disease prediction by recording the temporal patterns present in ECG dataset such as changes in oldpeak values that indicate heart stress and trends in exercise induced angina, improved by using LSTM [38].

c) **Multilayer Perceptrons (MLP)**: It is composed of fully connected layers which are designed to work best with data that are assembled in tabular form. The clinical and demographic dataset features like age, BMI, cholesterol, and stressors are efficiently evaluated by the use of MLP as it learn non-linear relationship exist among different features. It combines continuous and categorical data to produce accurate predictions, including blood pressure, cholesterol, and physical activity, to identify the risk of heart disease [39].

d) **Vision transformer (ViT):** It is initially designed for image data, where it function by partitioning the input data into patches and using self-attention techniques to find connections and relationship exists among the patches. ViT work best with global relationships when patch-based embedding are used to represent ECG time-series data or composite features. For instance, after encoding ECG signal patterns as input patches, ViTs can be used to analyse temporal dynamics across features such as Heart Rate Max and ST slope [40].
In this study, we have applied above models in order to improve the prediction accuracy of heart diseases by efficiently recognize structured and sequential data in combination with optimising feature extraction and selection techniques.

3.2.1 Training Layout of Models

Different training arrangements have been applied according to the model type and features present in dataset, to train machine learning model for ECG datasets. We have used PyTorch and TensorFlow/ Keras as the core framework to train and design the implementation of models. Theses framework is employed due to their versatility and resilience that make it efficient for investigation and application of ML models. The optimizer employed in conducting the training was Adam, due to its effectiveness with sparse gradients. Adam has the capability to combine the benefits of momentum-based optimisation and configurable learning rates, to enhances generalisation and speeds up convergence. All experiments were conducted using the default parameters ($\beta1 = 0.9$, $\beta2 = 0.999$, and $\varepsilon = 10^{-8}$). The initial learning that used was 0.001. For every 10 epochs, 0.1 factor was used to drop the learning rate in order to get finer updates in later training stages, that also improves stability and convergence. To balance the computing performance a batch size of 32 was utilized with gradient estimate stability. This size kept the GPU's memory requirements reasonable while allowing enough data each batch to compute steady gradients. A maximum of 50 epochs were used to train the models. However, using five-epoch patience, early stopping was used to track validation loss. Training was stopped to avoid overfitting and cut down on pointless calculations when validation loss did not enhance for five consecutive epochs. Improving model performance required addressing the dataset's class imbalance. SMOTE (Synthetic Minority Oversampling Technique) was used for augmentation to increase variety and lower the danger of overfitting, synthetic samples for the minority class were created in feature space. By using this strategy, the model was guaranteed to learn from both classes efficiently, producing predictions that were more evenly distributed. Since the goal variable is binary (1 = heart disease risk, 0 = no heart disease risk), the binary cross-entropy loss function was used. This loss function efficiently directs the model during optimisation by measuring the difference between true labels and anticipated probability. Google Colab, which gives access to the NVIDIA Tesla T4 GPU, whose GPU's 16 GB VRAM made it easier to process big datasets and train deep learning models and 12 GB RAM which provide enough memory to manage smaller datasets and successfully carry out data preparation procedures. This hardware setup was used for conducting all implementation. The model performances are compared on metrics parameters namely- accuracy, precision, recall, F1-Score, and training time. Figure 2; represent the flow chart of implementation of proposed methodology for feature selection method on ML models.

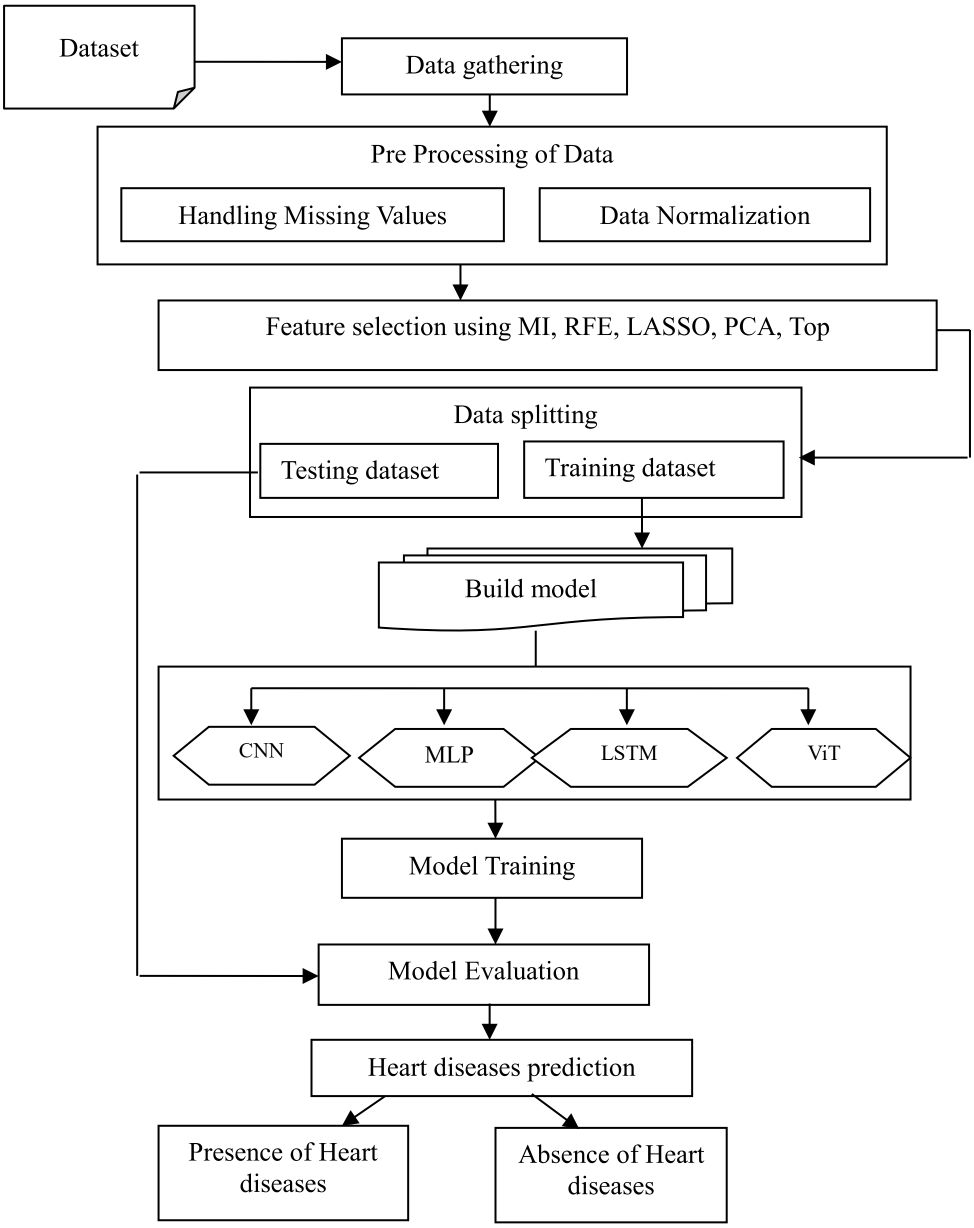

Figure 2: Proposed Methodology

3.2.2 Configurations Particular to a Model

The model specific configuration for model namely- CNN, LSTM, MLP, and ViT setups are incorporated below within Table 2.

Table 2: Training Configuration of ML Models

Model	Configuration
CNN	Three convolutional layers (Max Pooling, ReLU activation) and two dense layers make up the architecture. The size of the kernel is 3 × 3. Dropout rate: 0.5. There are [32, 64, 128] filters.
LSTM	The Architecture for LSTM consist of a dense layer comes after two LSTM layers (64 and 128 units). Tanh is the activation used for this model with Dropout rate: 0.3 and Sequence Length was 50 (truncated or padded as necessary).
MLP	Three dense layers with 128, 64, and 32 neurones make up the architecture. ReLU was utilized as activation function with Dropout rate of 0.4.
ViT	The architecture consists of patch size 16 × 16 with Eight Transformer layers and Embedding Dimension of 256. There are eight heads with 0.1 dropouts.

4. RESULT INTERPRETATION AND ANALYSIS

In this segment, we provide outcomes of the research that are obtained after the implementation of the proposed methodology. Based on the level of complexity of ML archetype, the different feature selection methods have selected ten different features from given 26 features set (provided in Table1), describe below with their indices as provided in Table 1:

Selected Features by RFE: [1, 2, 5, 9, 14, 16, 17, 21, 23, 24]
Selected Features by MI: [21, 7, 4, 11, 23, 17, 5, 2, 9, 1]
Selected Features by L1: [1, 2, 4, 5, 6, 7, 9, 10, 11, 12]
Selected Features by PCA: [15, 21, 5, 2, 11, 22, 19, 13, 9, 16]
Selected Features across all methods (Top): [20, 4, 1, 8, 10, 22, 16, 0, 15, 6]

By contrasting many machine learning models (CNN, MLP, ViT, and LSTM) used to predict heart disease using an ECG dataset, the Table 3 illustrates how feature selection strategies impact model performance and training time. Accuracy, precision, recall, F1 score, along with training time are among the evaluation metrics reported in each row for a model and feature selection method combination.

Table 3: Comparative Analysis of Model Performance

Model	Feature Selection Method	Accuracy	Precision	Recall	F1 Score	Training Time
CNN	**MI**	99.31%	99.32%	99.16%	99.30%	8.41
LSTM	**MI**	99.07%	99.08%	99.06%	99.07%	26.61
MLP	**RFE**	97.67%	97.71%	97.66%	97.67%	7.36
ViT	**RFE**	97.90%	97.95%	97.90%	97.90%	5.02

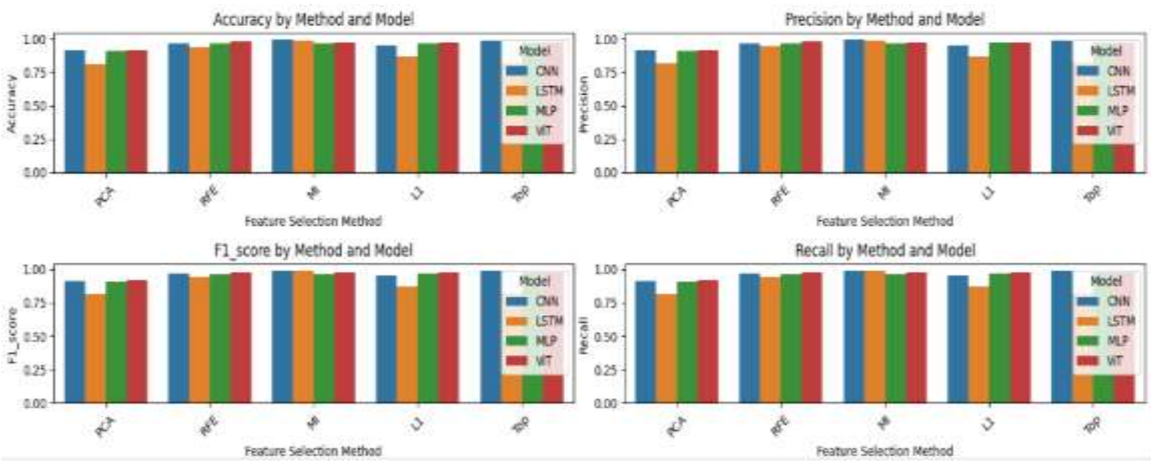

Figure 3: Performance plot of ML Model with Accuracy, Precision, F1 Score and Recall

Mutual Information (MI) is found best in respect of performance metrics namely- accuracy, recall, precision and F1 scores, and the most effective feature selection strategy for CNN and LSTM models (as shown in Figure 3). The MLP and ViT models perform better when using Recursive Feature Elimination (RFE), indicating that different models benefit from various feature selection techniques. At 99.31% accuracy, precision, recall, and F1 score, the CNN model using the MI feature selection strategy performed best across the board. When combined with MI, LSTM achieves 99.07% accuracy, as well as similarly excellent precision and recall, closely following CNN. For time-series classification tasks, CNN and LSTM perform well when combined with the MI feature selection strategy. RFE improves the performance of MLP and ViT, most likely because it improves generalisation following feature reduction. CNN and MI are suggested as the best combination for time-series data jobs in order to efficiently achieve top-tier performance. However, MLP and ViT in conjunction with RFE offer good substitutes when interpretability or other architectures are needed.

With the ten features consistently selected across methods Top, the CNN model performed the best across all evaluation metrics with a comparatively rapid training time of 9.86 seconds and an accuracy, precision, F1 score, and recall of 99.07%. While function together in combination with the feature selection approach-Top, CNN becomes the best model for the task at hand. Both the MLP and ViT models performed well, achieving high precision (~97.74%) and accuracy of 97.67%. However, ViT was more computationally efficient, taking only 4.62 seconds of training time as opposed to MLP's 7.13 seconds. With an accuracy of 82.05% and a longer training time (20.10 seconds), the LSTM model demonstrated poorer predictive performance, suggesting that, with the Top features used, it might not be the best option for this specific purpose.

5. CONCLUSION

This study investigated how various feature selection techniques affected the effectiveness of CNN, LSTM, MLP, and ViT machine learning models for classifying heart disease from ECG data. The study demonstrated how appropriate feature subsets enhance the productiveness, precision, and generalisation of ML archetype by utilising time-domain, frequency-domain, and nonlinear feature extraction strategies RFE, L1, PCA and MI. The outturn of this study present an idea about importance of feature selection for reducing data dimensionality, improving processing efficiency, and improving classification accuracy. CNN with Mutual Information (MI) outperformed the others models with the accuracy of 99.31% and short time of 8.41 seconds to train. Although it took a lot longer to train (26.61 seconds), LSTM with MI finished second with an accuracy of 99.07%. Because it was able to recognise the most instructive characteristics and record temporal relationships in sequential data, the MI technique was quite successful. Its intricate architecture and sequential processing probably contributed to its need for the greatest training duration (8.41 seconds). RFE, generated the best outcomes for fully connected networks like MLP and ViT. Despite extended training times, MLP and ViT achieved accuracies of 97.67% and 97.90%, respectively. For better performance and efficiency, CNN with MI is advised. When sequential linkages are especially complicated, alternative LSTM with MI can be employed. When choosing ViT and MLP architectures, RFE ought to be the primary feature selection technique, with ViT saved for more intricate situations that call for improved interpretability. By focussing on effective feature selection, model complexity, and training efficiency, this study provides useful model selection guidance for machine learning applications utilising imbalanced datasets.

Future studies should consider on integrating hybrid feature selection algorithms, using advanced Transformer-based designs for ECG time-series data. By resolving data imbalance concerns, validating models across many ECG datasets, and improving model interpretability, robustness and clinical application can be further improved.

REFERENCE

[1] Bani Hani, S.H. and Ahmad, M.M. (2023) ‘Machine-learning algorithms for ischemic heart disease prediction: A systematic review’, *Current Cardiology Reviews, 19*, pp. e090622205797. doi: 10.2174/1573403X19666220906222205797.

[2] Manji, R. A., Witt, J., Tappia, P. S., Jung, Y., Menkis, A. H., and Ramjiawan, B.(2013). ‘Cost–effectiveness analysis of rheumatic heart disease prevention strategies’, Expert Rev. Pharmacoecon. Outcomes Res., 13 (6), pp.715–724.

[3] Noroozi, Z., Orooji, A. and Erfannia, L. (2023). ‘Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction’, Scientific reports, 13, pp. 1–15. Available from: https://doi.org/10.1038/s41598-023-49962-w.

[4] Yekkala, I., Dixit, S. and Jabbar, M. (2017). ‘Prediction of heart disease using ensemble learning and Particle Swarm Optimization’, In 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon); IEEE (2017).

[5] Rautaharju, P. M., Surawicz, B., & Gettes, L. S. (2009). 'AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram', Journal of the American College of Cardiology, 53(11), 976-981.

[6] Badoni, P., Walia, R., and Mehra, R. (2024). 'Wearable IoT technology: Unveiling the smart hat', *Proceedings - 2024 1st International Conference on Intelligent Systems and Technologies for Emerging Markets (ISTEMS).* https://doi.org/10.1109/ISTEMS60181.2024.10560229

[7] Kligfield, P., Gettes, L. S., Bailey, J. J., Childers, R., Deal, B. J., Hancock, E. W., ... & Wagner, G. S. (2007). 'Recommendations for the standardization and interpretation of the electrocardiogram: part I: the electrocardiogram and its technology', Journal of the American College of Cardiology, 49(10), 1109-1127.

[8] Morris, F. (2008). 'ABC of Clinical Electrocardiography', Blackwell Pub: Oxford, UK.

[9] Clifford, G. D., Liu, C., Moody, B., Li, Q., Silva, I., Li-Pershing, Y., Behar, J., Johnson, A. E. W., Oster, J., Sapojnikov, M., Nemati, S., Scott, D. J., & Mark, R. G. (2017). 'AF classification from a short single lead ECG recording: The PhysioNet Computing in Cardiology Challenge 2017', Computing in Cardiology, 44, 1-4. https://doi.org/10.23919/CinC.2017.

[10] Najafi, A., Nemati, A., Ashrafzadeh, M. and Zolfani, S.H. (2023). 'Multiple-criteria decision making, feature selection, and deep learning: A golden triangle for heart disease identification', Engineering Applications of Artificial Intelligence,125. Available from: https://doi.org/10.1016/j.engappai.2023.106662.

[11] Satisha, C., Mehra, R., and Giri, M. (2023). Detection of various security attacks on IoT devices using machine learning. *2023 International Conference on Computational Intelligence and Networks (ICCINS)*. https://doi.org/10.1109/ICCINS58907.2023.10450058

[12] Badoni, P., Walia, R., & Mehra, R. (2024). 'Enhancing waste separation and management through IoT-based smart bin system', *Proceedings of the 2024 1st International Conference on Intelligent Systems and Technologies for Emerging Markets (ISTEMS)*. https://doi.org/10.1109/ISTEMS60181.2024.10560260

[13] Kumar, B., Soundararajan, R., Natesan, K.. and Santhi, R. M. (2023). 'Hybrid Feature Selection and Classifying Stages through Electrocardiogram (ECG) Signal for Heart Disease Prediction', Eng. Proc., MDPI, 59(126). https://doi.org/10.3390/ engproc2023059126

[14] Raj, S., & Ray, K. C. (2018). 'A personalized arrhythmia monitoring framework using wearable sensors and a deep learning network', Biomedical Signal Processing and Control, 44, 42-50. https://doi.org/10.1016/j.bspc.2018.04.001

[15] Kiranyaz, S., Ince, T., & Gabbouj, M. (2016). 'Real-time patient-specific ECG classification by 1-D convolutional neural networks', *IEEE Transactions on Biomedical Engineering, 63(3)*, 664-675. https://doi.org/10.1109/TBME.2015.2468589

[16] Golande, A. L. and Pavankumar, T. (2023). 'Optical electrocardiogram based heart disease prediction using hybrid deep learning', *Journal of Big Data* , 10(139). https://doi.org/10.1186/s40537-023-00820-6

[17] Li, X., Zhou, Y., Yu, J., & Wang, X. (2020). 'ECG signal classification using wavelet transform and deep convolutional neural networks', *Neural Computing and Applications, 32(10)*, 6849-6861. https://doi.org/10.1007/s00521-019-04144-1

[18] Kunadharaju, H. P. R., Sandhya, N., & Mehra, R. (2019). 'Multi-sensor image matching using super symmetric classifiers', *International Journal of Recent Technology and Engineering, 8*(2), 6161–6166. https://doi.org/10.35940/ijrte.B3764.078219

[19] Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V. and Nappi, M.(2021). 'Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques', *IEEE Access*, 9, pp. 39707–39716.

[20] Badoni, P., Walia, R., and Mehra, R. (2024). 'Wearable IoT technology: Unveiling the smart hat', *Proceedings - 2024 1st International Conference on Intelligent Systems and Technologies for Emerging Markets (ISTEMS).* https://doi.org/10.1109/ISTEMS60181.2024.10560229

[21] Vijay Krishnan, M. R., Mehra, R., & Kunadharaju, H. P. R. (2024). 'Design of neural network-based approaches for image processing', *Proceedings of the 2024 3rd International Conference on Electrical, Electronics, and Information Communication Technology (ICEEICT).* https://doi.org/10.1109/ICEEICT61591.2024.10718442

[22] Thiyagarajan, C. (2016). 'A survey on diabetes mellitus prediction using machine learning techniques', *Int. J. Appl. Eng*., 11, pp. 1810–1814.

[23] Ribeiro, A. H., Ribeiro, M. H., Paixão, G. M. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., Ferreira, M. P., Andersson, C. R., Macfarlane, P. W., Meira Jr, W., Schön, T. B., & Ribeiro, A. L. (2020). 'Automatic diagnosis of the 12-lead ECG using a deep neural network', *Nature Communications, 11*(1), 1760. https://doi.org/10.1038/s41467-020-15432-4

[24] Guyon, I., & Elisseeff, A. (2003). 'An introduction to variable and feature selection', *Journal of Machine Learning Research, 3*, 1157–1182.

[25] Acharya, U. R., Fujita, H., Lih, O. S., Adam, M., Tan, J. H. and Chua, C. K.. (2017). ' Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network' *Information Sciences, 405*, pp. 81-90. doi:10.1016/j.ins.2017.04.012

[26] Jadhav, S. M., Ghatol, A. A., & Holambe, R. S. (2010). 'Artificial neural network models based cardiac arrhythmia disease diagnosis from ECG signal data', *International Journal of Cardiology, 141(1)*, 1–12.

[27] Kachuee, M., Kiani, M. M., Mohammadzade, H., & Shabany, M. (2018). 'ECG heartbeat classification: A deep transferable representation', *IEEE Transactions on Biomedical Engineering, 66(5)*, 1196–1206. https://doi.org/10.1109/TBME.2018.2872652

[28] Yildirim, Ö., Talo, M., Ay, B., Baloglu, U. B., Aydin, G., and Acharya, U. R. (2018). 'Automated arrhythmia detection using deep learning techniques: A review', *Computers in Biology and Medicine*, 96, pp. 189-202.

[29] Sadhukhan, P., Roy, K., and Biswas, S. (2020). 'Comparative Analysis of Feature Selection Techniques for ECG-based Myocardial Infarction Detection', *Journal of Biomedical Signal Processing and Control*, 58, 101869.

[30] Peker, M., Demir, D., and Yildirim, B. (2019). 'The impact of feature selection on neural networks and support vector machines for arrhythmia classification', *Journal of Medical Systems, 43*(6), 178. https://doi.org/10.1007/s10916-019-1357-0.

[31] Mincholé, A. et al. (2019). 'Impact of Feature Selection on Machine Learning Models for ECG Signal Classification', *Journal of Medical Systems*, 43(4), 92. https://doi.org/10.1007/s10916-019-1324-7.

[32] Zhao, Y., Liu, L., and Zhang, Y. (2021). 'Feature selection for heart disease detection using machine learning: A focus on heart rate variability', *Journal of Medical Systems*, 45(8), pp. 1-12. https://doi.org/10.1007/s10916-021-01762-4.

[33] Kiranyaz, S., Ince, T. and Gabbouj, M. (2018). 'Personalized Monitoring of ECG Signals With Hybrid Feature Selection and Deep CNN Models', *IEEE Transactions on Biomedical Engineering*, 65(4), pp. 1015-1023. DOI: 10.1109/TBME.2017.2754295.

[34] Liang, J., Wang, J., and Zhang, Z. (2020). 'Pre-selecting features for deep learning models: Improving convergence speed and reducing overfitting in ECG-based heart disease detection', *IEEE Transactions on Biomedical Engineering*, 67(4), pp.1025-1034.

[35] Choi, H. et al. (2020). 'Enhancing ECG-Based Myocardial Infarction Detection Using Random Forest and Feature Importance Metrics', *Journal of Medical Systems*, Vol. 44, pp. 1-12.

[36] Chen, Y., Wang, X., Li, Z. and Zhang, J. (2021). 'Hybrid Feature Selection for Enhancing ECG-Based Deep Neural Networks in Heart Abnormality Detection,' *Journal of Medical Signal Processing*, 45(3), pp. 245–260.

[37] Lu, R. (2019). 'Malware detection with LSTM using opcode language'. *arXiv preprint arXiv:1906.04593*. https://arxiv.org/abs/1906.04593

[38] Sharma, R., Kamble, S. S., Gunasekaran, A., Kumar, V., & Kumar, A. (2020). 'A systematic literature review on machine learning applications for sustainable agriculture supply chain performance', *Computers & Operations Research, 119*, 104926. https://doi.org/10.1016/j.cor.2020.104926

[39] Aeinfar, V., Mazdarani, H., Deregeh, F., Hayati, M., & Payandeh, M. (2009). 'Multilayer Perceptron Neural Network with supervised training method for diagnosis and predicting blood disorder and cancer', In *2009 IEEE International Symposium on Industrial Electronics* (pp. 2075-2080). IEEE. https://doi.org/10.1109/ISIE.2009.5213842

[40] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). 'BERT: Pre-training of deep bidirectional transformers for language understanding'. *arXiv preprint arXiv:1810.04805*. https://arxiv.org/abs/1810.04805