

## RESEARCH ARTICLE

# Heart Disease Prediction Using Novel Ensemble and Blending Based Cardiovascular Disease Detection Networks: EnsCVDD-Net and BICVDD-Net

HIRA KHAN<sup>1</sup>, NADEEM JAVAID<sup>1,2,3</sup>, TARIQ BASHIR<sup>1</sup>, MARIAM AKBAR<sup>2</sup>, NABIL ALRAJEH<sup>4</sup>, AND SHERAZ ASLAM<sup>5</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, COMSATS University Islamabad, Islamabad 44000, Pakistan

<sup>2</sup>Department of Computer Science, COMSATS University Islamabad, Islamabad 44000, Pakistan

<sup>3</sup>International Graduate School of Artificial Intelligence, National Yunlin University of Science and Technology, Douliu, Yunlin 64002, Taiwan

<sup>4</sup>Department of Biomedical Technology, College of Applied Medical Sciences, King Saud University, Riyadh 11633, Saudi Arabia

<sup>5</sup>Department of Electrical Engineering, Computer Engineering and Informatics (EECEI), Cyprus University of Technology, Limassol 3036, Cyprus

Corresponding author: Nadeem Javaid (nadeemjavaidqau@gmail.com)

This project is funded by Researchers Supporting Project number (RSPD2024R648), King Saud University, Riyadh, Saudi Arabia.

**ABSTRACT** Cardiovascular Diseases (CVDs) have emerged as a significant physiological condition, being a primary contributor to mortality. Timely and precise diagnosis of heart disease is crucial to safeguard patients from additional harm. Recent studies show that the usage of data driven approaches, such as Deep Learning (DL) and Machine Learning (ML) techniques, in the field of medical science is highly useful in accurately diagnosing heart disease in less time. However, statistical learning and traditional ML approaches require feature engineering to generate robust and effective features from data, which are then used in the prediction models. In the case of large complex data, both processes pose many challenges. Whereas, DL techniques are capable of learning features automatically from the data and are effective at handling large and intricate datasets while outperforming the ML models. This study focuses on the accurate prediction of CVDs, considering the patient's health and socio-economic conditions while mitigating the challenges presented by imbalanced data. The Adaptive Synthetic Sampling Technique is used for data balancing, while the Point Biserial Correlation Coefficient is used as a feature selection technique. In this study, two DL models, Ensemble based Cardiovascular Disease Detection Network (EnsCVDD-Net) and Blending based Cardiovascular Disease Detection Network (BICVDD-Net), are proposed for accurate prediction and classification of CVDs. EnsCVDD-Net is made by applying an ensemble technique to LeNet and Gated Recurrent Unit (GRU), and BICVDD-Net is made by blending LeNet, GRU and Multilayer Perceptron. SHapley Additive exPlanations is used to provide a clear understanding of the influence different factors have on CVD diagnosis. The network's performance is evaluated on the basis of various performance metrics. The results indicate that the EnsCVDD-Net outperforms all base models with 88% accuracy, 88% F1-score, 91% precision, 85% recall, and 777s execution time. Similarly, with 91% accuracy, 91% F1-score, 96% precision, 86% recall, and 247s execution time, BICVDD-Net outperforms the state-of-the-art DL models. To validate the model's results, 10-Fold Cross Validation is employed. An eXplainable Artificial Intelligence technique, SHapley Additive exPlanation is employed to know the features contribution in model's predictions.

**INDEX TERMS** Cardiovascular disease detection, deep learning, heart disease, LeNet, gated recurrent unit, multilayer perceptron, eXplainable artificial intelligence, SHapley additive exPlanation.

## I. INTRODUCTION

The associate editor coordinating the review of this manuscript and approving it for publication was Juan Wang<sup>1</sup>.

Cardiovascular Diseases (CVDs) are considered one of the major causes of mortality around the world [1]. All conditions

that affect the functioning of the heart and its vessels, and the ways in which the blood is pumped and is compelled to flow throughout the body come under the umbrella of CVDs. The heart is one of the most important organs in the human body. Hence, heart diseases are potentially fatal or in some cases leave severe long-term impairments on its patients [2]. During recent decades, CVDs have increasingly afflicted the global population. According to the World Health Organization's (WHO) survey, heart diseases account for 32% of the global deaths, of which 85% are because of heart attacks and strokes [1]. The increase in the number of heart disease patients is imposing a significant burden on hospitals worldwide [3]. Therefore, early detection of CVDs is critical so that the disease management through medication and counseling can start as early as possible.

Many risk factors play an important role in causing CVDs. These include unhealthy diets, a lack of physical activity, obesity, diabetes, and the use of tobacco and alcohol [4]. These daily life choices can cause conditions like high Blood Pressure (BP), high cholesterol, and insulin resistance, which in turn increase the possibility of developing CVDs. Late detection of CVDs results in invasive procedures like angiography or bypass surgeries. These procedures not only cause discomfort for patients but also create a burden on healthcare facilities. Heart disease prediction is affected by parameters including age, sex, family history of CVDs, BP, cholesterol level and diabetes. Besides, it is a difficult task for doctors to identify the early symptoms of heart disease because several factors like hypertension, arrhythmia, hyperlipidemia, etc. are involved [5].

This strong need for timely detection of CVDs and reducing the death toll caused by heart diseases has led to the use of intelligent CVD detection systems. Since the introduction of Artificial Intelligence (AI), it has played a significant role in improving the quality of life, especially by helping in the early detection of diseases and saving human lives [6]. The intelligent CVD detection systems not only help with timely detection but also offer a better alternative to the traditional orangemethods by reducing the chance of human error. These systems perform CVD prediction using patients' health record attributes, by looking for the correlations between these attributes [7].

In the recent years, AI models are becoming popular for early and accurate disease prediction. The quality of a dataset plays an important role in achieving better performance using an AI model. The healthcare datasets are a collection of features like age group, gender, BP, cholesterol levels, and other health related features of different individuals. These features are analyzed by the AI models for detecting patterns that relate to higher chances of heart disease. The diversity of a dataset affect the accuracy of a model. The more diverse the dataset will be, the better the model will work on different types of patients. However, as the datasets get diverse, they increase in size and ML models may not perform well on big data [8].

This created the need for deeper Neural Networks (NNs) that could handle large datasets. Deep Learning (DL) models are the evolution of Machine Learning (ML) models which can not only be trained on bigger datasets but also offer other advantages like automated feature detection, handling unorganized data, extensibility and improved performance than ML models [9]. Deep Neural Networks (DNNs) are designed such that they behave like the human brain's NN, and this makes it possible for it to learn complex data and make accurate predictions for the healthcare industry. Owing to the exceptional performance of DL models in disease prediction, in this study, two DL models have been proposed for the accurate prediction of CVDs.

Predicting heart disease is a challenging task. Medical datasets are highly imbalanced which makes disease prediction complicated. Also, with the increase in the size of the medical datasets correct selection of model has become crucial for timely successful disease prediction. Data imbalance in disease prediction can significantly impact the performance of predictive models. In the literature, this issue is tackled using Synthetic Minority Oversampling TEchnique (SMOTE). However, SMOTE, due to class overlapping, may not effectively address the complex nature of disease data and generates noisy synthetic data samples [10], [11], [12]. Data augmentation techniques such as Mixup, and Time-series Generative Adversarial Networks (TimeGAN) have also been used for data balancing. However, Mixup and TimeGAN, though effective in generating synthetic data, struggle to replicate the intricate patterns and domain-specific characteristics inherent in disease-related sequential data, resulting in inadequate representations [13]. Model selection poses challenges, such as the use of Convolutional Neural Network (CNN) and Locally Weighted Random Forest (LWRF) for disease prediction, being resource-intensive and sensitive to hyper-parameters [14]. Bidirectional Long-Short Term Memory (Bi-LSTM) introduces prediction delays unsuitable for dynamic disease predictions [12]. Most of the studies have utilized ML algorithms for disease prediction [11], [15]. However, ML algorithms cannot handle noisy data as effectively as DL methods do. Furthermore, there is a lack of research that extends the prediction results beyond performance metrics [16]. All the abbreviations are provided in Table 15 (see Appendix).

## A. MAIN CONTRIBUTIONS

In this paper, we have made the following key contributions.

- For precise CVD prediction, we have proposed a DL-based model, Ensemble based Cardiovascular Disease Detection Network (EnsCVDD-Net), which combines the predictions of LeNet and Gated Recurrent Unit (GRU) in an ensemble approach.
- We have proposed a blending-based model of LeNet, GRU, and Multilayer Perceptron (MLP) called Blending based Cardiovascular Disease Detection Network (BICVDD-Net) for effective CVD detection.

- Adaptive Synthetic (ADASYN) Sampling Technique is used for data balancing, to generate synthetic samples that can more accurately represent the complex nature of disease data.
- Point Biserial Correlation Coefficient (PBCC) is utilized as a feature selection method to identify the most significant features for heart disease prediction within the dataset and enhance computational efficiency of the proposed models.
- 10-Fold Cross Validation (10-FCV) has been applied to verify the performance of the proposed EnsCVDD-Net and BICVDD-Net models.
- To learn the contribution of features in predicting heart disease, an eXplainable AI (XAI) technique, SHapley Additive exPlanations (SHAP) has been employed on EnsCVDD-Net and BICVDD-Net.

The remainder of this study is organized into the subsequent sections: Section II provides a detailed discussion of prior work conducted in the realm of heart disease prediction. Section III outlines the methodology of the proposed DL frameworks, encompassing specifics of the dataset, pre-processing, and model development. Section IV shows the experimental outcomes of the proposed models. Lastly, Section V provides the conclusion and future research direction of our study.

## II. RELATED WORK

CVD is a physiological illness and recently there has been a lot of study on the use of ML and DL algorithms to predict physiological disease. In [17], optimization of Artificial Neural Networks (ANN) using Genetic Algorithm (GA) has been investigated for CVD prediction. The dataset contains 12 features of 70,000 instances collected from medical records. Some other ML algorithms like Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) are also applied to the dataset. GA-ANN outperforms other algorithms with 73.43% accuracy.

George et al. [18] utilize data assimilation methods, such as the Kalman filter, to make predictions based on sparse, non-stationary Electronic Health Record (EHR) data in the Intensive Care Unit (ICU). The newly developed Constrained Ensemble Kalman Filter (CEKF) is combined with ML methods to enhance the accuracy of inference with sparse clinical data. The paper also highlights the limitations of Kalman filtering methods in clinical settings and identifies new problems that need to be addressed to make inference feasible using realistic clinical data. The proposed methodology reduces the Mean Squared Error (MSE) on the sparse clinical data.

In [19], the authors propose a stacking method for heart disease prediction. It utilize Gradient Boosting-based Sequential Feature Selection (GBSFS) for feature selection from heart disease dataset. Then the stacking model of various ML algorithms, such as DT, RF, Multilayer Perceptron (MLP), SVM, Extra Tree (ET), Gradient Boosting (GB), Linear

Regression (LR), and KNN, are trained on the dataset with the features selected from GBSFS. For the testing phase, unseen data points are fed to the base learners, and their individual predictions are made. Finally, the meta learner combined these predictions and gave the final prediction of heart disease for the new data points. By achieving a test accuracy of 98.78%, this study suggested that the combination of feature selection and ensemble learning can be effective for heart disease prediction.

In [20], the authors predicts death, heart failure, myocardial infarction, or stroke within a variable prediction horizon of 1 to 5 years. The paper employed a Recurrent Neural Network (RNN) with binary cross-entropy of the 25 outputs to analyze large datasets of administrative claims. Hyperparameter tuning has been performed in two steps using the Adaptive Moment estimation (ADAM) algorithm. The model is designed to minimize data preparation costs and demonstrate satisfactory performance in predicting major adverse cardiovascular events at all prediction horizons, with Area under the Receiver Operating Characteristic Curve (AUCROC) ranging from 0.812 to 0.792 and C-index ranging from 0.802.

Hymavathi et al. [21] discuss the serious threat to global health that heart disease poses, as it can take many different forms that impact the cardiovascular system. The authors provide a ML based ensemble model that uses meta-features to diagnose cardiac disease. Weighted average ensemble, adaptive boosting, RF with feature bagging, negative correlation learning, and ensemble of small models are the five strategies combined in the suggested study. For comparison, traditional ML techniques are also used. Experimental results demonstrate the improved performance of the ensemble model for accurately predicting the heart disease. The integration of meta-features enhanced the model's interpretability and adaptability, highlighting its potential for practical healthcare applications.

In [22], the authors review the application of DL in cardiology, focusing on diagnostic imaging, medical services, and biological NNs. It has highlighted various DL models, such as CNN, Auto-Encoders (AEs) using U-net architecture, RNN with Long Short Term Memory (LSTM) and GRU. The paper proposes that DNN with Rectified Linear Unit (ReLU) as compared to logistic sigmoid and hyperbolic tangent (tanh) function yielded better performance. To cater the imbalance in the dataset, SMOTE has been studied. The paper evaluates different DL models like convolutional, DNNs and ANNs with hidden layers in cascading patterns to process nonlinear datasets along with ML algorithms. The hybrid model combining convolutional and RNNs has yielded higher performance with an accuracy of 98.77%.

Zheng and Hu [24] proposed Disease Progression via Longitudinal Data Fusion (DPLDF) method which has a lower Mean Squared Error (MSE) as compared to LR with Least Absolute Shrinkage Selection Operator (LASSO) regularization models. The predictive performance of the models is quantified by MSE where DPLDF has 3.212 MSE

TABLE 1. Related Work Summary.

Methodology	Limitations	Evaluation Metrics
ANN with GA, an optimization strategy in the NN is employed [17]	Optimization of the performance of ANNs for CVD prediction	Accuracy
Data assimilation (DA) with CEKF is proposed [18]	Data sparsity and physiological prediction with clinical data	MSE
GBSFS to extract significant features from a heart disease dataset to train DT, RF, MLP, SVM, ET, GB, LR, KNN, and a stacking model [19]	Difficulty in handling the complexity and variability of medical datasets and lack of a robust feature selection technique	Test accuracy, Cross-Validation score (CV score), precision, recall, and F1-score
A RNN model is used to predict major adverse cardiovascular events in diabetic patients [20]	Accessing and processing large volumes of data generated from the disease management processes of diabetic patients is challenging	AUCROC, C-Index
Creation of end-to-end models that learn from the data without requiring external steps or prior knowledge [22]	Most segmentation methods are not end-to-end and rely on preprocessing, handcrafted features, or non-differentiable methods	Accuracy, sensitivity, specificity, F1-score, precision
Feature selection using $X^2$ statistical model and optimally configured DNN is used [23]	Underfitting and overfitting problems in heart disease prediction	Accuracy
Disease Progression via DPLDF formulation is proposed [24]	Precise prediction of the potential disease progression based on the knowledge in the EHR data	Average Correlation Coefficient (R), MSE
Various ML algorithms including SVM, NB, RF and KNN are used [25]	Address sentiment analysis challenges in informal text, particularly due to morphological complexities and dialectal variations	Precision, recall, F1-measure, accuracy
Advanced SSL technique is proposed for working on two modalities corresponding to time-series data, and spectrogram in time-frequency domain [26]	Lack of exploitation of ECG characteristics in frequency domain	AUCROC, accuracy, F1-score
Contextualized embeddings like BERT and ELMo work better than traditional, non-contextualized embeddings, in RNN architectures [27]	The majority of detection techniques concentrated on basic metadata like followers, post text, and account activity patterns. Without taking into account posting times, languages, or semantic coherence, they were readily misled by keyword stuffing.	Accuracy, precision, recall and F1-score
Pretrained Med-BERT on a large and diverse EHR dataset using Masked Language Model (LM) and Prolonged Length Of Stay (Prolonged LOS) tasks is proposed [28]	Small and incomplete training data for DL models	Average AUCROC values and Standard Deviation (SD)
ImageCHD, the first dataset for CHD classification, is introduced along with a baseline framework for automatic CHD classification using a cutting-edge segmentation method [29]	Lack of dataset for Chronic Heart Disease (CHD) classification	Accuracy
The deep CNN-LSTM model with wavelet transform and median filtering is proposed to capture and learn hierarchical and temporal features for improved disease recognition from time series data [30]	DL models lack in capturing and propagating temporal dependencies among the physiological signals	Accuracy, precision, recall, F1-score
C-BiLSTM algorithm is proposed to accurately predict heart diseases by capturing temporal patient data variations [32]	Suboptimal accuracy of traditional heart disease prediction methods	Accuracy
A hybrid CNN and LSTM method is proposed and compared with various ML algorithms [33]	ML techniques fail to achieve high accuracy for heart disease prediction	Accuracy
A BiLSTM-GRU model, a hybrid RNN model hyper-parameter tuned using RSCV technique is proposed predicting Coronary Heart Disease [34]	Small data sets for classification, predominant use of ML over potentially superior DL, and incomplete evaluation metrics	Sensitivity, Specificity, Negative Predicted Value, Positive Predicted Value, F1-score, Accuracy

TABLE 1. (Continued.) Related Work Summary.

The study used embeddings from LLMs to transform posts into vector representations. Various ML models, including SVM, RF, XGBoost, KNN, and NNs, are used [37]	Previous works often encountered inconsistent results and limitations in handling diverse linguistic expressions	Accuracy, precision, recall
Prediction of the occurrence of chronic diseases by employing the ML technique, KNN, DT, and DL employing with Adam serving as an optimizer is proposed [38]	Poor accuracy of prior techniques	Accuracy
A new ensemble QMBC technique for identifying patients diagnosed with heart disease and those who are not diagnosed is used [40]	Individual models suffer from overfitting, biased predictions, or incomplete representation of the underlying patterns in the data	Accuracy, precision, specificity, recall, F1-score
Development of CNN model supported by TensorFlow Lite for micro-controllers and optimizing the TinyCES for low-power, resource-constrained environments in embedded systems [42]	The setup demands significant resources for communication and computation, which can be problematic in terms of efficiency and resource usage	Accuracy, model size, network bandwidth usage, size of recorded ECG data

which is less than LR with LASSO regularization. However, this methodology requires extensive data preprocessing to handle inconsistencies and missing values, which can be a significant limitation, particularly in cases where patient data is sporadic or incomplete.

Alfreihaat et al. [25] propose a framework for sentiment analysis that considers both emojis and text features. Emoji Sentiment Lexicon (Emo-SL) assign sentiment scores to frequently used emojis allowing the framework to capture the emotional weight emojis carry in online communication. The study incorporate sentiment scores from Emo-SL alongside sentiment analysis from existing text-based lexicons. Various ML models, including SVM, Naive Bayes (NB), RF, and KNN, are used in the research. The combined approach, using Emo-SL derived features with ML techniques, improve the accuracy by 26.7% and give the F1-score of 89%.

The research paper [26] on the classification of arrhythmias in multi-lead Electrocardiogram (ECG) data using multi-modal approaches incorporates Self-Supervised Learning (SSL) algorithms to analyze unlabeled ECG data. The network proposed in this study utilizes SSL and consists of two modules corresponding to pre-stream (self-knowledge distillation techniques employed in the absence of labeled data) and down-stream tasks (trained on labeled data, a gate fusion mechanism is integrated to merge information from different modalities, blending features from both time-series data and spectrograms). The model achieved an AUCROC of 96%, demonstrating its effectiveness on the given dataset.

Social media spam detection, focused on DL architectures with RNNs to capture long term dependencies in text is proposed in [27]. Instead of relying on static word embeddings, Alshattnawi, S. et al. incorporated contextualized embeddings like Bidirectional Encoder Representations from Transformers (BERT) and Embeddings from Language Model (ELMo). These models considered the surrounding words to understand a word’s meaning in a better manner. This approach led to an improvement (over 10-15%) in

identifying spam in datasets of both Twitter and YouTube. The study highlighted the potential of contextualized language models for improving social media security by developing more effective defense mechanisms against evolving spam threats.

In the development of Medical BERT (Med-BERT) [28], a pre-trained model using large-scale EHR data has been used. The paper demonstrates the application of DL and Natural Language Processing (NLP) in disease prediction. The models include variations of GRU, Bidirectional GRU (BiGRU), Reverse Time Attention Mechanism (RETAIN), and Med-BERT, tested on three different datasets: Diabetes Heart Failure cohort (DHF-Cerner), Pancreatic Cancer cohort (PaCa-Cerner, and PaCa-Truven). The combination of Bi-GRU with Med-BERT consistently shows high performance across all three datasets, with accuracies of 83.15%, 82.63%, and 79.17% for DHF-Cerner, PaCa-Cerner, and PaCa-Truven, respectively.

The creation of a specialized 3D Computed Tomography (CT) Image dataset for Classifying Congenital Heart Diseases (ImageCHD) [29], highlights the importance of high-quality, disease-specific data in medical diagnostics. The methodology focuses on advanced image processing techniques to analyze and classify complex congenital heart conditions from 3D CT scans. The base method provides an accuracy of 81.9%.

In [6], AI-based learning techniques are reviewed for their role in disease diagnosis and post-operative life expectancy prediction. This paper highlights a specific case study on a data-driven cervical cancer prediction model. A deep CNN-LSTM network for recognizing diseases using multivariate physiological signals has been utilized. LSTM achieved an accuracy of 95.7%.

In [30], the authors used an approach for disease recognition by treating physiological signal data as a time series classification problem in ML. This perspective is particularly beneficial for diseases where physiological signals play a key

role in diagnosis. The 4CNN+2LSTM provides an F1-score of 0.88.

Bouchareb et al. [31] describe a system that enhances prediction accuracy using multiple classifiers. This method exemplifies the trend towards more sophisticated, integrated computational models in healthcare, capable of handling diverse data types for more accurate disease prediction. LR with Principal Component Analysis (PCA) outperforms with 98% to 100% accuracy and with 95% to 98% accuracy with Chest X-ray (CXR) images for feature extraction.

The authors in [32] presents a heart disease prediction model that utilizes a Cluster-based bidirectional LSTM (C-BiLSTM) algorithm. This approach combined the advantages of clustering for data segmentation with the capabilities of BiLSTM. The C-BiLSTM model is good in acquiring temporal dependencies and variations in patient data. The accuracy score of the proposed system is 94% for real data and 92% for University of California Irvine (UCI) data.

Sudha and Kumar [33] in their research suggested a hybrid CNN and LSTM based system for CVD prediction. This combined the power of feature extraction of CNN with the capability of handling sequential data efficiently that LSTMs possess to make a robust framework for analyzing medical data. This hybrid model's accuracy is 89%, sensitivity is 81%, and specificity is 93%.

Sharma et al. made a hybrid DNN learning model for predicting Coronary Heart Disease, and used Randomized Search Cross Validation (RSCV) optimization [34] with 98% accuracy. Naskath et al. [35] provide detailed comparison of different DL algorithms used in DNNs, focusing on MLP, Self-Organizing Maps (SOMs), and Deep Belief Networks (DBNs) by highlighting their applicability in various healthcare scenarios, including heart disease prediction. The response and performance matrix scores are above 90% for these algorithms.

Ogundepo et al.'s [36] research performs a performance analysis of supervised classification models on heart disease prediction. This study underscores the critical role of model selection in predictive accuracy, examining various supervised learning models to determine their efficacy in heart disease prediction. SVM achieves an 87% accuracy, 88% specificity, 93% AUC, 85% sensitivity, 36% log loss, 85% precision and 85% F1-score.

In [37], the authors explore the potential of using Large Language Models (LLMs) and ML algorithms for predictive analytics in mental health. The study aim to capture the semantic meaning and details of language, in order to analyze social media posts. Following that, different ML models, including SVM, RF, Extreme Gradient Boosting (XGBoost), KNN, and NNs are trained with these embeddings to determine whether or not posts are suggestive of mental health conditions like stress disorders. SVM outperforms other ML techniques and yielded the highest accuracy of 83% in classifying posts related to stress.

The paper [38] focused on the use of DL and ML techniques, such as KNN, DT, and DL with ReLU and sigmoid activation functions, to predict chronic diseases in patients. Praveenkumar, S. et al. discuss the prediction of patients' incurable diseases using a DL approach, indicating the expanding role of AI in diagnosing complex medical conditions. The DL approaches produce a superior accuracy, which is around 97.9%. The study by Shukur and Mijwil [39] involves ML techniques in heart disease diagnosis, providing a performance analysis of various models including LR, RF, ANN, SVM, and KNN. SVM provides the highest accuracy of 96%.

Kapila et al. introduce a novel Quine McCluskey Binary Classifier (QMBC) for heart disease prediction [40]. The proposed QMBC (Anova with PCA) provides an accuracy of 98.36%, proposed QMBC (Chi-Square with PCA) provides an accuracy of 99.92% on CVD Dataset, and proposed QMBC (Anova with PCA) on heart disease comprehensive dataset results in 98.31% accuracy.

Dhanka et al. present uses supervised algorithms for coronary artery heart disease detection, contributing to the field of expert systems [41]. After parametric optimization, the accuracy score for LR is 87.78% and XGBoost is 91%.

Kim et al. [42] developed Tiny ML based Classification for ECG monitoring Embedded Systems (TinyCES). The framework used Tiny ML (TinyML) for classifying ECG data directly on the monitoring device. It enabled the device to perform the classification using a small, optimized ML model. This reduced the amount of data needed to be transmitted resulting in improved battery life. Two large ECG datasets: the Massachusetts Institute of Technology-Beth Israel Hospital Arrhythmia Database and the Physikalisch-Technische Bundesanstalt Diagnostic ECG Database, are used to train model. With an approximately 97% detection ratio, TinyCES yielded high performance in monitoring ECG.

Wang et al. [43] uses a Cloud-RF (C-RF) algorithm for risk assessment of coronary heart disease, blending cloud computing and ML, for healthcare applications and achieves an accuracy score of 85%. In summary, these studies demonstrate a significant focus on the application of advanced ML and DL techniques, optimization algorithms, and data preprocessing methods for early-stage and accurate prediction of heart disease. The integration of various computational approaches, including ensemble methods, NNs, and feature selection techniques, illustrates the evolving landscape of data-driven healthcare solutions. Summary of the prior studies is shown in Table 1.

### III. PROPOSED SYSTEM MODELS

This section outlines the overall flow of this study. Two ensemble techniques, EnsCVDD-Net and BICVDD-Net, are proposed in our work. Before being fed into the DL models, The dataset must undergo preprocessing. The dataset used is cleaned, i.e., it does not contain any missing values or

outliers. However, it is highly imbalanced. Class 0 contains more instances than Class 1. Thus, we have employed the ADASYN balancing approach to balance the data. In both proposed approaches, LeNet and GRU are used. It leverages the strengths of both CNN for feature extraction and RNN for considering not just the current input, but also a memory of past inputs.

### A. DATASET

For this study “Heart Disease Health Indicators Dataset,” is used. It is a patient health record dataset gathered for heart disease prediction [44]. Centers for Disease Control (CDC) made the Behavioural Risk Factor Surveillance System (BRFSS) for conducting a yearly health survey that produced this dataset. This extensive dataset contains the annual health data of more than 400,000 Americans, which makes it one of the largest health survey datasets in the world. It has been running since 1948.

In this dataset, each row is composed of detailed information of a heart patient. The Heart Disease Health Indicator dataset is organized in a tabular manner. This dataset includes 22 features and 253,680 instances, which assists in the binary categorization of heart disease. These features include age, smoking habits, body mass index, alcohol intake, history of stroke, and physical and mental health status, and are gathered via individual replies or direct participant queries. Heat map [45] is used to visualize the correlation between each feature in the dataset, as shown in Fig. 1.

Beyond missing values and outliers, the dataset has already been cleansed. Class 0 accounts for 91% of occurrences in the dataset, whereas class 1 accounts for just 9% occurrence. It is crucial to recognize this significant imbalance in the dataset. As such, preprocessing measures like class balancing and feature selection are necessary steps to guarantee the robustness and dependability of the trained Ensemble and Blending CardioNets in practical settings. Heart disease prediction research and development can benefit from the public availability of this dataset on Kaggle.

### B. DATA BALANCING

One of the crucial steps in preprocessing is data balancing. A large difference in the number of instances between two classes indicates an imbalance in the data. Two classes can be found in a dataset: the minority class, which has fewer occurrences, and the majority class, which has more. An imbalanced dataset can lead to a poor and biased model performance.

It is a significant difficulty for real-time databases. Two main data balancing techniques are used to overcome this problem of data imbalance: under-sampling and over-sampling. In under-sampling technique, the majority class instances are decreased to align with the minority class, but this may result in the loss of important information. On the other hand, in over-sampling technique, the instances in the minority class are increased by creating synthetic samples.

There are several techniques for balancing data. To avoid information loss, it is important to perform this process carefully. The dataset used in this study is highly imbalanced with 229,787 participants in class 0 (no heart disease) and only 23,893 participants in class 1 (heart disease). It shows that a small proportion of the individuals in the dataset have heart disease. To address this imbalance, the ADASYN data balancing technique has been used in this study.

### C. ADAPTIVE SYNTHETIC SAMPLING TECHNIQUE

ADASYN is an effective technique for handling imbalanced datasets. ADASYN analyses the distribution of the minority class and assigns weights to each data point, unlike simple oversampling such as Random Over Sampling (ROS), which simply duplicates the existing minority data. These weights reflect the “difficulty” of learning from each point; harder-to-classify data points receive higher weights, signifying their need for more synthetic neighbors. Based on these weights, ADASYN then generates new data points, effectively focusing its efforts on the most challenging areas of the minority class [46].

ADASYN ensures that the DL model will learn the characteristics of the minority class accurately by decreasing the bias towards the majority class. Another of its advantages is that it concentrates on the data samples that are difficult to classify and this helps the model in efficiently distinguishing the class boundaries in complex regions. Unlike its counterparts that replicate the existing data points for data balancing, ADASYN generates new synthetic data samples that helps in preventing overfitting of the model.

All these advantages make ADASYN an effective method for overcoming class imbalance and improving the performance of the model on the minority classes, especially in the domain of medical diagnosis, and anomaly recognition [47]. Fig. 2 shows that by utilizing the strengths of ADASYN balancing technique, the severe imbalance within the Heart Disease Health Indicator dataset is tackled.

### D. FEATURE SELECTION

Another crucial DL preprocessing step is feature selection or dimension reduction. It is a simple and effectual method to remove redundant and irrelevant data. It improves learning accuracy and interpretability of the models in decision making process and reduces computational time. In this paper, PBCC technique is used for feature selection.

A statistical method known as PBCC is used to measure the correlation between a continuous and binary variable [48]. This technique is mostly used when one of the variables is binary, i.e., it has only two possible outcomes. In this study, binary variable is the target variable having two possible outcomes: healthy individuals and heart diseased patients. For feature selection, PBCC can be employed to evaluate the direction and intensity of each feature’s correlation with a binary target variable. The most significant attributes in predicting the binary outcome can be determined by

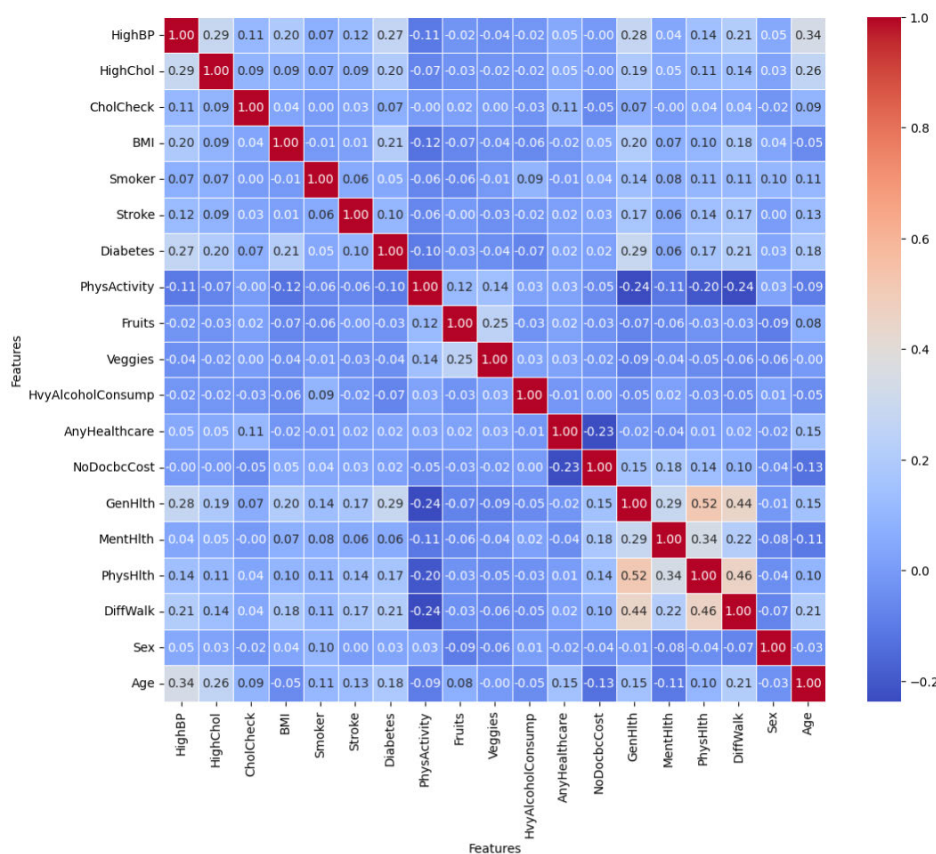


FIGURE 1. Correlation heatmap of heart disease health indicator dataset.

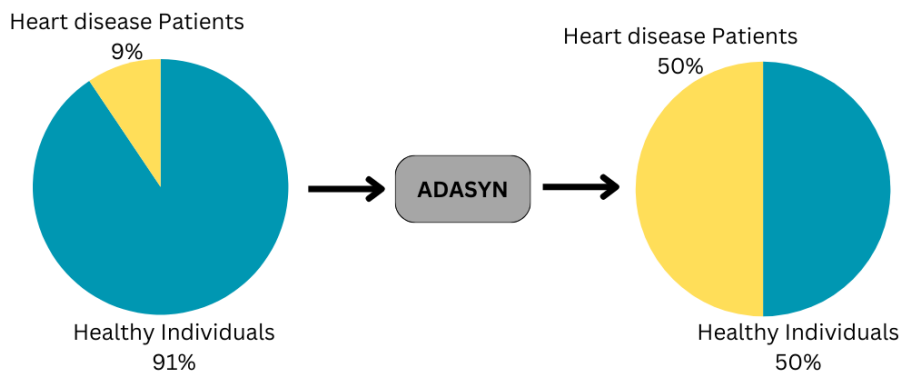


FIGURE 2. Data balancing through adaptive synthetic sampling technique.

computing this coefficient for every feature. Features with higher absolute values of the PBCC are considered to have greater influence.

**IV. DESCRIPTION OF ENSEMBLE AND BLENDING BASED CARDIOVASCULAR DISEASE DETECTION NETWORKS (ENSCVDD-NET AND BLCVDD-NET)**

In DL, an ensemble technique involves combining the predictions of multiple NNs to improve overall performance. It provides advantages such as reduced variance and lower

generalization error as it overcomes the individual model’s limitations. These techniques are used for various purposes, from finance to healthcare, like detecting heart disease, theft detection in smart grids, image recognition and so on. In ensemble techniques, the models are trained and their predictions are combined using different methods like averaging, voting and stacking. Some of the popular ensemble techniques are hybrid, bagging, boosting and stacking.

In this research, two ensemble techniques, EnsCVDD-Net and BICVDD-Net, have been proposed as shown in Fig. 3.



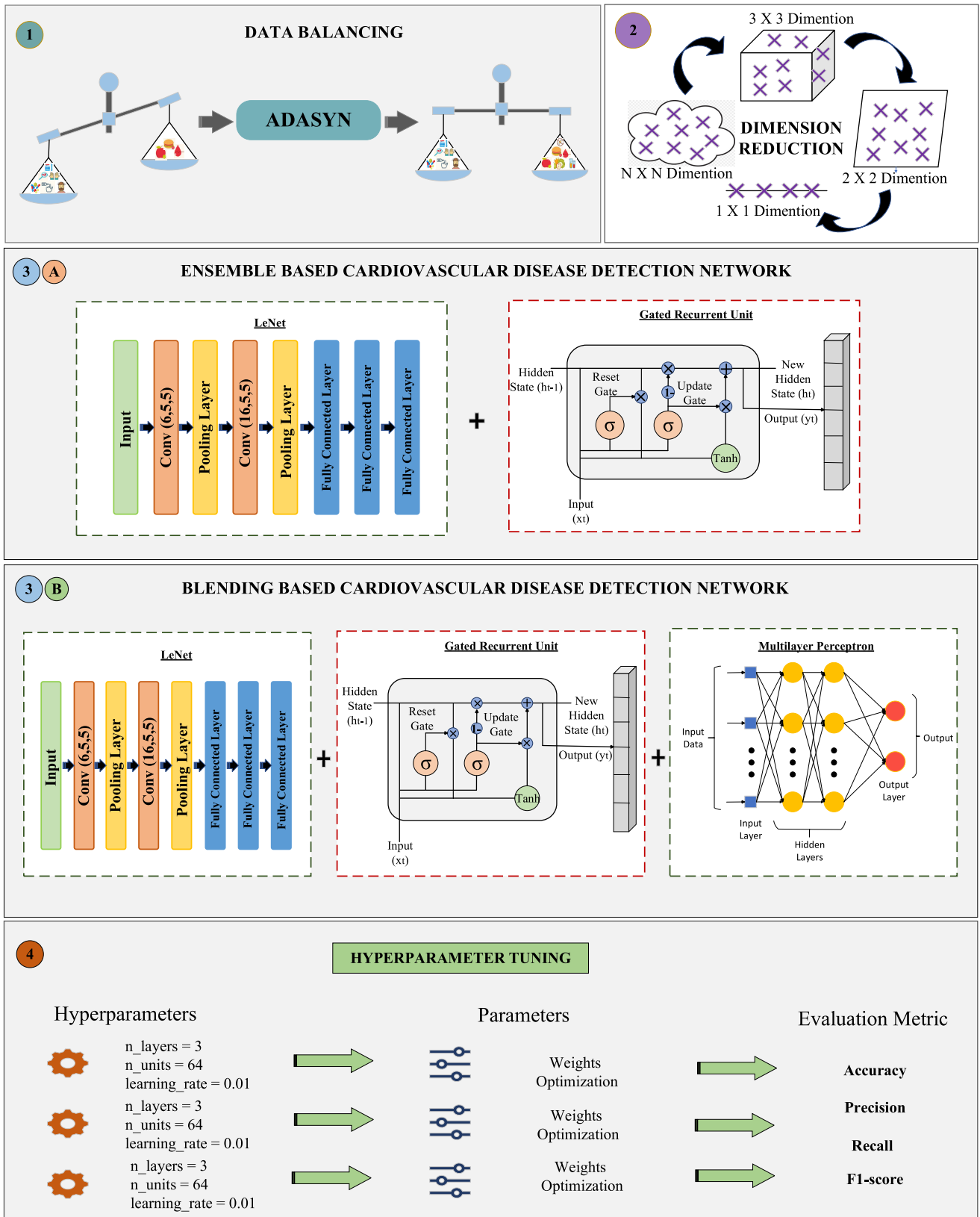


FIGURE 3. Proposed system model for heart disease prediction.

EnsCVDD-Net is a hybrid combination of LeNet and GRU, while BICVDD-Net is a blending of LeNet and GRU as base models with MLP as the meta-learner layer.

### A. ENSEMBLE BASED CARDIOVASCULAR DISEASE DETECTION NETWORK

A hybrid model combines different DL models to form a single predictive model. It minimizes the drawbacks of each model while maximizing its strengths. By merging multiple models, the hybrid model delivers superior predictive performance compared to using a single model alone.

Sequential hybrid modeling or parallel hybrid modeling are the two approaches that can be used to create a hybrid model. In a sequential hybrid model, the output of the first base model is given as an input to the second base model, resulting in a single output. When using a parallel hybrid model, each model functions on its own. Based on the same input, both models give predictions. The outputs of both models are then concatenated to create a single output, which makes them hybrid [49]. A hybrid model requires careful consideration of several factors, such as the values of the hyper-parameters and the training strategies.

Our proposed EnsCVDD-Net model comprises of two DL models, i.e., LeNet and GRU. The combined use of both models enhances the overall performance. Both the models received the same input. In the case of LeNet, the input passes through convolutional layers, then max-pooling layers and fully connected layers. By adjusting the kernel sizes and strides of the convolutional layers accordingly, the output is obtained through features. While GRU processes the input data sequentially, using the recurrent nature of the GRU layers. The output of both models is concatenated and passed through the last hybrid layer for final prediction. The workflow of the EnsCVDD-Net is shown in Algorithm 1.

The GRU is known for its effectiveness in sequential data processing. In the proposed model, EnsCVDD-Net, the GRU is comprised of two layers, with 17 neurons and a total of 929 parameters. LeNet is one of the pioneers in CNN, which contains convolutional, pooling, and fully connected layers to learn intricate features from the data. For EnsCVDD-Net, the LeNet architecture is modified to have 12 layers, with 6929 parameters and 173 neurons. These modifications to the LeNet architecture help in accurate heart disease detection. The hybrid of these models results in a comprehensive architecture, EnsCVDD-Net, consisting of 16 layers, 7857 parameters, and 189 neurons. It aimed at improving the accuracy and reliability of heart disease predictions by leveraging the strengths of both GRU and LeNet architectures.

### B. BLENDING BASED CARDIOVASCULAR DISEASE DETECTION NETWORK

The blending technique combines the predictions from multiple individual models to create a final prediction. This

### Algorithm 1 Ensemble Based Cardiovascular Disease Detection Network for Heart Disease Prediction

**Require:** Dataset  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

- 1: Split the data into features  $X_{\text{train}}$  and labels  $Y_{\text{train}}$
- 2: **Input:**
- 3: Input  $X$
- 4: Initialize LeNet and GRU models with their respective architectures and parameters.
- 5: Pass input  $X$  through LeNet model:
- 6: Let  $Y_{\text{LeNet}} = \text{LeNet}(X)$
- 7: Pass input  $X$  through GRU model:
- 8: Let  $Y_{\text{GRU}} = \text{GRU}(X)$
- 9: Concatenate outputs of LeNet and GRU models:
- 10: Let  $Y_{\text{concat}} = \text{Concatenate}(Y_{\text{LeNet}}, Y_{\text{GRU}})$
- 11: Pass concatenated output through a fully connected layer:
- 12:  $Y_{\text{output}} = \text{Dense}(Y_{\text{concat}})$
- 13: Apply sigmoid activation to obtain final output:
- 14:  $\hat{Y} = \text{Sigmoid}(Y_{\text{output}})$
- 15: **Output:** Final heart disease prediction.

reduces the chances of the model getting overfitted and enhances the overall performance of the blending based model. In this study, two base models, LeNet and GRU, and a meta-model, MLP, have been blended together to make BICVDD-Net. For blending, the dataset is split into three parts: training, validation and testing. The base models, LeNet and GRU, are trained on the training set. After training, they generate predictions on the validation set. These predictions become the new features for training the meta-model, MLP. At the end, MLP learns on these new features and accurately generates the final prediction on unseen data, as shown in Algorithm 2.

## V. DEEP LEARNING TECHNIQUES

The proposed ensemble models, EnsCVDD-Net and BICVDD-Net, are composed of different DL models. This section discusses in detail each deep model used in the proposed ensemble models.

### A. LENET

LeNet is one of the initial CNN architectures that was proposed by LeCun, Y. et al. in the late 1990s. LeNet 5's architecture is made up of 7 layers: 3 convolutional, 2 pooling and 2 fully connected layers. As the name of the model suggests, out of these 7 layers, 5 of them are learnable [50]. The architecture of LeNet 5 is shown in Fig. 4. It is primarily designed for handwritten digit recognition. It is a well-known architecture because of its simplicity and efficient pattern identification abilities. Due to this, in this research, LeNet is used for heart disease detection.

In LeNet, the convolutional layers play a significant role in identifying hidden patterns and intricate relationships within the data. Following the convolutional layers, the pooling layers are used that reduce the dimensionality of the feature

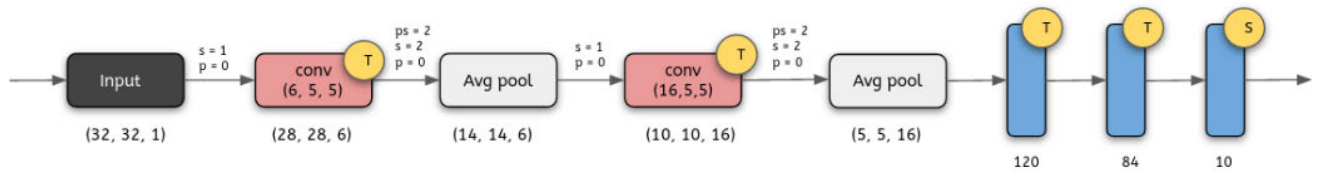


FIGURE 4. LeNet architecture.

### Algorithm 2 Blending Based Cardiovascular Disease Detection Network for Heart Disease Prediction

**Require:** Dataset  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

- 1: Split  $D$  into training set  $D_{train}$ , validation set  $D_{valid}$ , and testing set  $D_{test}$
- 2: **Input:**
- 3: Input  $X$
- 4: Initialize LeNet, GRU and MLP models with their respective architectures and parameters
- 5: Train LeNet model  $M_{LeNet}$  on  $D_{train}$  by passing input  $X$  through LeNet model
- 6: Obtain LeNet predictions  $P_{LeNet}$  on  $D_{valid}$ :  
 $P_{LeNet} = LeNet(X)$
- 7: Train GRU model  $M_{GRU}$  on  $D_{train}$  by passing input  $X$  through GRU model
- 8: Obtain GRU predictions  $P_{GRU}$  on  $D_{valid}$ :  
 $P_{GRU} = GRU(X)$
- 9: Combine  $P_{LeNet}$  and  $P_{GRU}$  into  $X_{valid\_combined}$
- 10: Train MLP model  $M_{MLP}$  on  $X_{valid\_combined}$  with labels  $y_{valid}$
- 11: Generate final predictions using  $M_{MLP}$  on  $X_{test}$
- 12: **Output:** Final heart disease prediction.

maps without losing the most important data in order to make the network less complex. The first convolutional layer performs 1D convolution on 1D heart disease data using 5 kernels and 32 neurons, followed by a pooling layer. The second convolutional layer, equipped with 8 neurons and the same kernel size, further refines the feature maps. The third convolutional layer applies a  $1 \times 1$  convolution with 16 neurons.

Next, there are 3 fully connected layers with 32, 16 and 4 neurons, respectively. The first 5 learnable layers use a tanh activation function, and in the final dense layer, a sigmoid activation function is used. This final layer outputs the probability of the presence of heart disease. This combination of activation functions is employed to ensure a balance between computational efficiency and non-linearity within the network.

LeNet is an effective model and is simple to implement [51]. This makes LeNet a popular choice in the field of DL. The alternating convolutional and pooling layers in LeNet's architecture followed by fully connected layers, provide a robust framework for heart disease prediction tasks. Algorithm 3 shows the workflow of LeNet model.

### Algorithm 3 Working of LeNet

**Require:** Dataset  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

- 1: Split the data into features  $X_{train}$  and labels  $Y_{train}$
- 2: **Input:**
- 3: Input  $X$
- 4: Initialize the weights  $w$  and biases  $b$  for each layer  $L$ .
- 5: Pass the input through multiple convolutional layers followed by max-pooling layers.
- 6: **Convolutional layer:**  
 $Y_1 = \text{Tanh} \left( \sum_{i=1}^{K_1} (X * W_{1i} + b_{1i}) \right)$
- 7: **Max-Pooling layer:**  
 $Y_2 = \text{Max.Pooling}(Y_1, P_1)$
- 8: **Fully-connect layer:** After the convolutional and pooling layers, three fully connected layers are used for classification.
- 9:  $Y_3 = \text{Flatten}(Y_2)$
- 10: **Output layer:** The final layer produces class probabilities using sigmoid activation function.
- 11:  $\hat{Y} = \text{Sigmoid}(Y_3 \cdot W_2 + b_2)$
- 12: **Output:** Predicted class label.

### B. GATED RECURRENT UNIT

As a kind of RNN, GRU is intended to simplify the gating mechanism and minimize the number of parameters, hence improving the computational performance and simplifying training. The gated procedures are also used to manage and regulate the information flow in the NNs. It is developed to address some shortcomings of RNN, such as vanishing gradients, while maintaining the ability to capture sequential dependencies effectively. GRU enables capturing long-term dependencies from massive sequential data without excluding information from the previous portion of the sequence of data.

GRU uses a reset gate and an update gate and a hidden state to transfer information from one cell to another, as shown in Fig. 5. The reset gate chooses how much of the previous information to lose, whereas the update gate chooses how much of the previous information to save. The information from earlier time steps is stored in the hidden state, which functions as a kind of memory. It evolves over time as GRU processes sequential input [52].

The previous hidden state ( $h_{t-1}$ ) and current input ( $x_t$ ) are combined and passed through a reset gate.

$$r_t = \sigma [W(r)x_t + U(r)h_{t-1}] \quad (1)$$

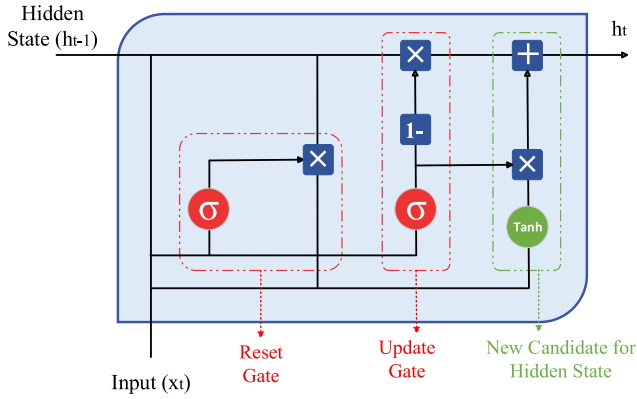


FIGURE 5. Architecture of gated recurrent unit.

While the upgrade gate updates the hidden state with new information.

$$z_t = \sigma[W(z)x_t + U(z)h_{t-1}] \quad (2)$$

The GRU computes a candidate hidden state that blends the previous hidden state with new information from the current input.

$$\tilde{h}_t = \tanh(W_h x_t + U_h h_{t-1}) \quad (3)$$

$$h_t = z_t h_{t-1} + (1 - z_t) \tilde{h}_t \quad (4)$$

GRU is effective in applications where sequential data processing is required in the field of DL [53]. The GRU is less complex in architecture with fewer parameters when compared to LSTM and other variants of RNN with gating mechanisms. When working with limited resources, this improves the GRU's computational efficiency.

#### Algorithm 4 Working of Gated Recurrent Unit

**Require:** Dataset  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

- 1: Split data into features  $X_{train}$  and labels  $Y_{train}$
- 2: **Input:**
- 3: Input  $x_t$ , Previous hidden state  $h_{t-1}$ , Weight matrices  $W_r, W_z, W_h, U_r, U_z, U_h$
- 4: Compute reset gate:  $r_t = \sigma(W_r x_t + U_r h_{t-1})$
- 5: Compute update gate:  $z_t = \sigma(W_z x_t + U_z h_{t-1})$
- 6: Compute candidate hidden state:  $\tilde{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1}))$
- 7: Compute new hidden state:  $h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t$
- 8: **Output layer:**
- 9: Final prediction.

### C. MULTILAYER PERCEPTRON

A type of ANN called MLP is made up of several layers of interconnected neurons. The data flow of a feed-forward NN is unidirectional, meaning it originates at the input layer, moves through hidden levels, and ultimately arrives at the output layer. Each node-to-node connection has a weight, and

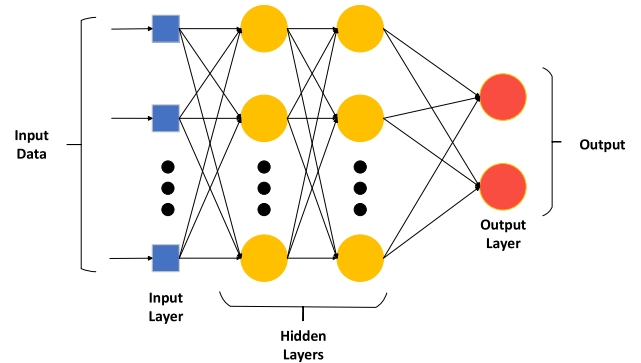


FIGURE 6. Architecture of Multilayer Perceptron.

the network learns to adjust these weights during training to make accurate predictions [54].

An input layer, a hidden layer or layers, and an output layer are the three primary types of layers that make up an MLP. Fig. 6 describes the basic architecture of MLP. The input layer consists of nodes that represent the features of the input data. For the input features  $x_1, x_2, x_3, \dots, x_n$ , the output of the input layer is the same as the input values, and it is denoted as  $a^0$ :

$$a^0 = [x_1, x_2, x_3, \dots, x_n] \quad (5)$$

MLP comprises of one or multiple hidden layers sandwiched between the input and output layers. Every node within a hidden layer is linked to all nodes in its adjacent layers, both preceding and succeeding [55]. The output for each node in the hidden layer is determined by applying an activation function to the weighted sum of its inputs.

Common activation functions include the sigmoid function, tanh or ReLU. The output layer produces the final output of the network. The calculation is similar to that of the hidden layers, using the weighted sum and an activation function.

$$z_k^L = \sum_{j=1}^{N^{L-1}} w_{kj}^L \cdot a_j^{L-1} + b_k^L \quad (6)$$

$$a_k^L = \sigma(z_k^L) \quad (7)$$

where  $k$  indexes the output nodes,  $N^{L-1}$  is the number of nodes in the previous layer,  $w_{kj}^L$  is the weight of the connection,  $a_j^{L-1}$  is the output of the  $j$ -th node in the previous layer,  $b_k^L$  is the bias term and  $\sigma$  is the activation function. The architecture of MLP is trained using a supervised learning approach, adjusting the weights and biases to minimize a loss function. This is typically done using backpropagation and gradient descent to reduce error. In Algorithm 5,  $x$  is the input,  $w$  represents weights,  $h$  represents hidden layers, and  $l$  is the hidden layers count.

### D. PROPOSED MODELS' PERFORMANCE EVALUATION

The confusion matrix is a tool that is used for evaluating the performance of these proposed models, EnsCVDD-Net,

**Algorithm 5** Working of Multilayer Perceptron**Require:** Dataset  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 

- 1: Split data into features  $X_{train}$  and labels  $Y_{train}$
- 2: Initialize weights  $w^{(l)}$  and biases  $b^{(l)}$  for each layer  $l$
- 3: **Input layer**
- 4:  $a^{(0)} = x$
- 5: **for**  $l = 1$  to  $L$  **do**
- 6:     **Hidden layer**
- 7:      $z^{(l)} = w^{(l)}a^{(l-1)} + b^{(l)}$
- 8:      $a^{(l)} = \sigma(z^{(l)})$
- 9: **end for**
- 10: **Output layer**
- 11: Final prediction.

**TABLE 2.** Confusion matrix.

	Actual: Heart disease Patient	Actual: Healthy Patient
Predicted: Heart disease Patient	TP	FP
Predicted: Healthy Patient	FN	TN

BICVDD-Net and the individual models. The confusion matrix is made of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) as shown in Table 2, and it is an effectual method for measuring the performance metrics like accuracy, precision, recall and F1-score.

**E. EVALUATION METRICS**

This section will provide a detailed description of the performance metrics that are used to assess how well the proposed models perform.

**Accuracy** is one of the most instinctive performance evaluation measures that shows the percentage score of the accurate predictions made by the model in predicting heart disease, as compared to the total observations. Equation 8 is used for calculating the accuracy.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

**Precision** can be seen as the indication of exactness and is the ratio of TP observations to all positive observations predicted by the model. In terms of heart disease prediction, it shows the accuracy of positive predictions the model has made. High value of precision reflects a low rate of false positives. It is calculated by equation 9:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

**Recall** shows the ratio of TP that the model correctly predicts against all the observations in actual class. In disease

prediction, it describes the model's ability to identify all positive cases. It is calculated by equation 10:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

**F1-score** is a model evaluation performance metric that combines precision and recall using harmonic mean. It focuses on the class-wise performance rather than the overall performance of the model. It includes both FP and FN and is considered more useful than accuracy. It is calculated by equation 11:

$$F1\text{-score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (11)$$

**F. K-FOLD CROSS VALIDATION**

$k$ -FCV technique is used to evaluate the performance of our model on unseen data by repeatedly splitting the data into  $k$  folds and using them for training and validation [56]. The dataset is split into equal sized  $k$  folds. Then, in order to prevent overfitting, the dataset is trained and evaluated  $k$  times. For  $k$ -FCV, we split the dataset into three sets: training, testing, and validation. Every time a different fold is selected for validation [57]. The model is trained over  $k$  iterations. In each iteration,  $k-1$  folds are used for training and the remaining folds are used for validation. Accuracy score along with F1-score, precision, recall, and execution time are obtained after training and testing. In our study, we have taken  $k = 10$ , resulting in 10-FCV. 10-FCV increases the model's efficacy by yielding highly reliable outputs [58].

**G. ESTIMATING UNCERTAINTY: 95% CONFIDENCE INTERVAL FOR ENSEMBLE AND BLENDING BASED CARDIOVASCULAR DISEASE DETECTION NETWORKS RESULTS**

After applying 10-FCV on our proposed models, we get 10 values for each metric. We calculate 95% Confidence Interval (CI) from those values, to estimate the range within which the true performance of our DL models likely falls. For  $n$  number of values, the point estimate, which is the mean value,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  and the standard deviation  $s$  are calculated (equation 12).

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (12)$$

We used the standard deviation of the performance metric in order to calculate the margin error, which is given by

$$ME = t_{\alpha/2, df} \times \frac{s}{\sqrt{n}} \quad (13)$$

where  $t_{\alpha/2, df}$  is the critical value from the t-distribution value and  $df = n - 1$  degrees of freedom. The value of  $t_{\alpha/2, df}$  is equal to 2.201 for determining 95% CI with small  $df$ . Adding and subtracting this value from the point estimate, we obtain the upper and lower bounds of the 95% CI for our

deep models as shown in equation 14.

$$95\%CI = \bar{x} \pm ME \tag{14}$$

By the inclusion of CI, we gained a better understanding of our models' generalizability and the potential variability in their performance.

**H. RESULTS OF ENSEMBLE BASED CARDIOVASCULAR DISEASE DETECTION NETWORK**

EnsCVDD-Net is the hybrid of LeNet and GRU models. Table 3 and Table 4 provides the details of EnsCVDD-Net architecture and the hyper-parameters utilized in it, respectively. The provided Tables 5, 6, 7 and 8 present the results of an EnsCVDD-Net across 5, 10, 20 and 30 epochs, respectively. It showcases the performance of three different models: LeNet, GRU and EnsCVDD-Net. These tables show various evaluation metrics, including accuracy, precision, recall, F1-score and execution times at different training epochs. For each performance metrics, the tables include 95% CI for the EnsCVDD-Net model at different epochs of 10-FCV. The bold value is the mean or average value while the values enclosed in square brackets are the CI values calculated after applying 10-FCV on EnsCVDD-Net.

The EnsCVDD-Net, when equipped with PBCC, exhibits a notable improvement in accuracy as compared to LeNet and GRU at the 5-epochs mark (Table 5). Despite the improvements, the execution time of the EnsCVDD-Net remains reasonable, making it a practical choice.

The performance of EnsCVDD-Net continues to improve at 10 epochs, with the highest accuracy among the three models, as shown in Table 6. Precision, recall, and F1-score are well-balanced, indicating the model's reliability. While the execution time is higher, it still remains manageable.

Table 7 shows that the EnsCVDD-Net is still leading in terms of accuracy at 20 epochs, performing better than both LeNet and GRU. Precision, recall, and F1-score values also suggest that the PBCC technique contributes positively to the model's performance. The execution time increases with the increase in the number of epochs but is still within the acceptable limits.

From Table 8 it can be seen that with the increase in the number of epochs, the performance of EnsCVDD-Net increases, while still having the highest accuracy and competitive precision, recall, and F1-score. The execution time, as expected, is larger than that observed for earlier epochs, but is reasonable considering the increase in performance.

The inclusion of PBCC appears to enhance the EnsCVDD-Net's performance, particularly in terms of accuracy, precision, and recall. This technique likely aids in better feature selection and classification, resulting in improved overall results. However, it is important to consider the trade-off between performance gains and execution time, as the EnsCVDD-Net with PBCC may require more computational resources as compared to the original model.

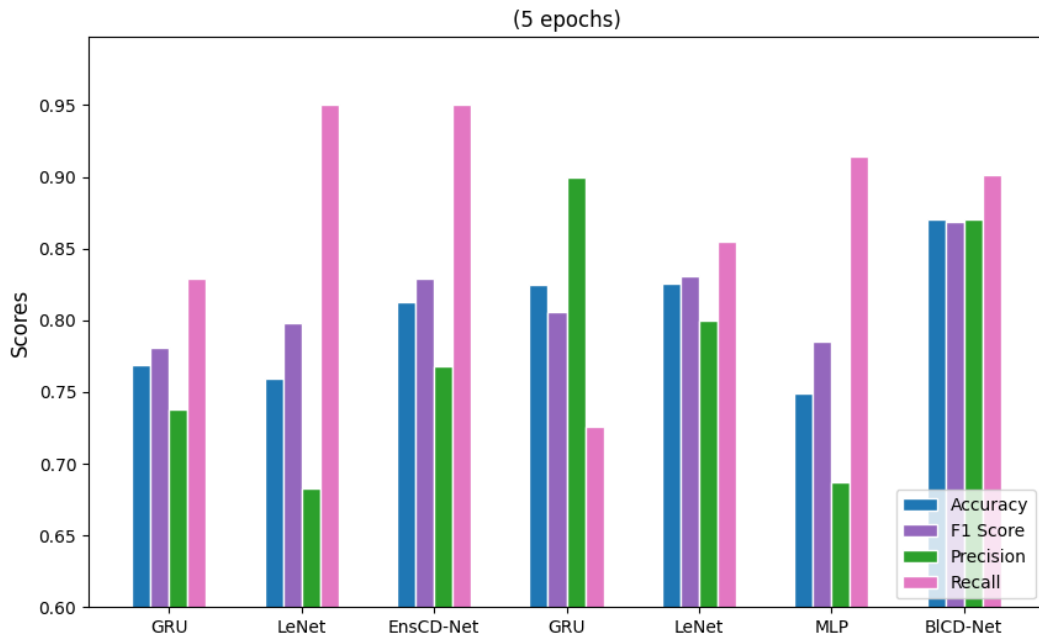
**TABLE 3. Architectures of deep learning models and ensemble based cardiovascular disease detection network.**

Models	Architecture
LeNet	Conv 1D layer(Number of neurons=32, kernel_size=5, activation='tanh') MaxPooling layer(padding = 'same') Conv 1D layer(Number of neurons=8, kernel_size=5, activation='sigmoid') MaxPooling layer(padding = 'same') Conv 1D layer(Number of neurons=16, kernel_size=5, activation='tanh') Dense layer(Number of neurons=64, activation function='linear') Dense layer(Number of neurons=32, activation='tanh') Dense layer(Number of neurons=16, activation='linear') Dense(Number of neurons=4, activation='tanh') Dense layer(num_classes, activation function='sigmoid')
GRU	GRU(Number of neurons=16, activation='tanh', return_sequences=False) Dense layer(Number of neurons=1, activation function='sigmoid')
EnsCVDD-Net	Conv 1D layer(Number of neurons=32, kernel_size=5, activation='tanh') MaxPooling layer(padding='same') Conv 1D layer(Number of neurons=8, kernel_size=5, activation='sigmoid') MaxPooling layer(padding='same') Conv 1D layer(Number of neuron=16, kernel_size=5, activation='tanh') MaxPooling layer(padding='same') Dense layer(Number of neuron=64, activation='tanh') Dense layer(Number of neuron=32, activation='tanh') Dense layer(Number of neurons=16, activation='tanh') Dense layer(Number of neurons=4, activation='tanh') GRU(Number of neurons=16, activation='tanh', return_sequences=False) Dense layer(Number of neurons=2, activation function='sigmoid')

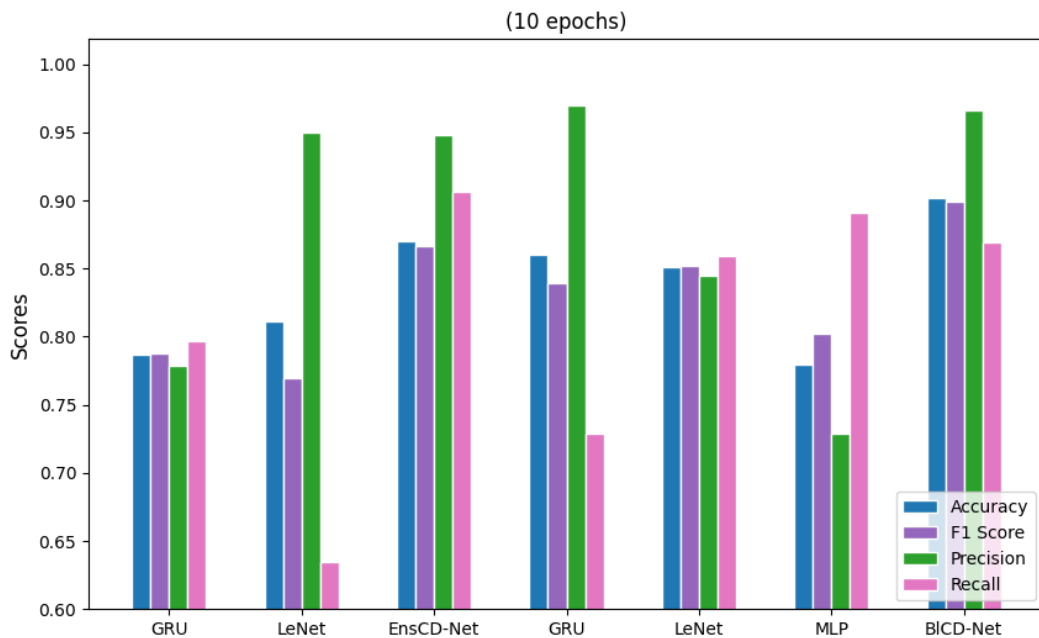
**I. RESULTS OF BLENDING BASED CARDIOVASCULAR DISEASE DETECTION NETWORK**

The BICVDD-Net combines two base models: LeNet, GRU, and a meta-model, MLP. The details of the architecture of BICVDD-Net and the hyper-parameters utilized in it are shown in Tables 9 and 10 respectively. At the 5-epochs, the BICVDD-Net demonstrates competitive accuracy of 87%, precision of 87.03%, recall of 86.9%, and F1-score of 86.9%, as shown in Table 11. It stands out with the highest performance metrics, while LeNet and GRU also provide strong contributions. Execution times for all models are reasonable, with the BICVDD-Net being the fastest.

At 10 epochs, the BICVDD-Net continues to deliver strong performance across all metrics, with an accuracy of 89.4% and an F1-score of 89.3% (Table 12). Execution times are consistent with Table 11. It also maintains its competitive advantage at 20 epochs, with an accuracy of 91.1% and



**FIGURE 7.** Simulation representation of ensemble and blending based cardiovascular disease detection networks on 5 epochs for heart disease prediction.

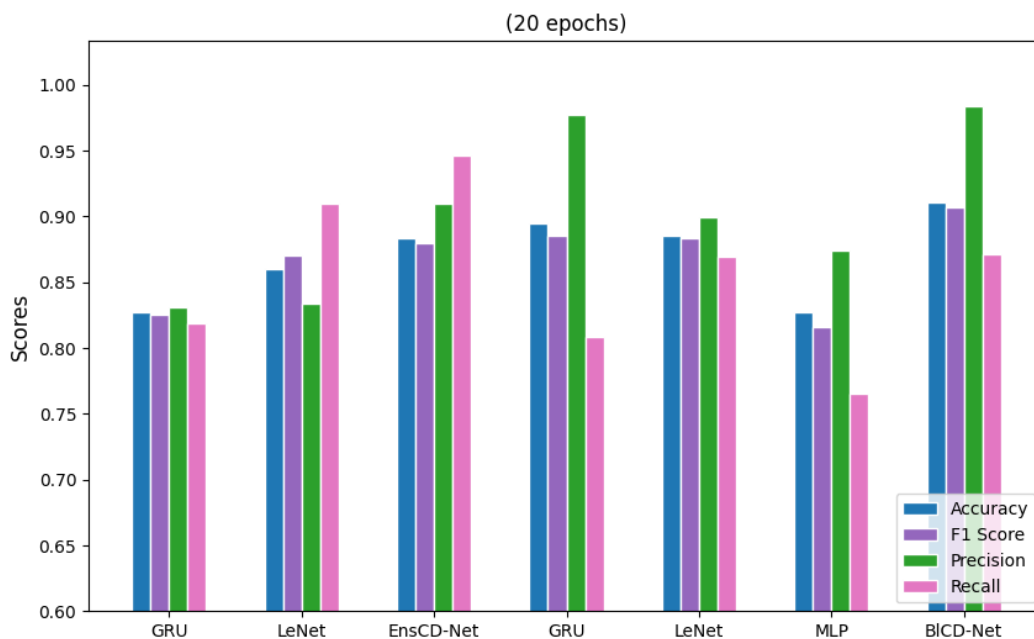


**FIGURE 8.** Simulation representation of ensemble and blending based cardiovascular disease detection networks on 10 epochs for heart disease prediction.

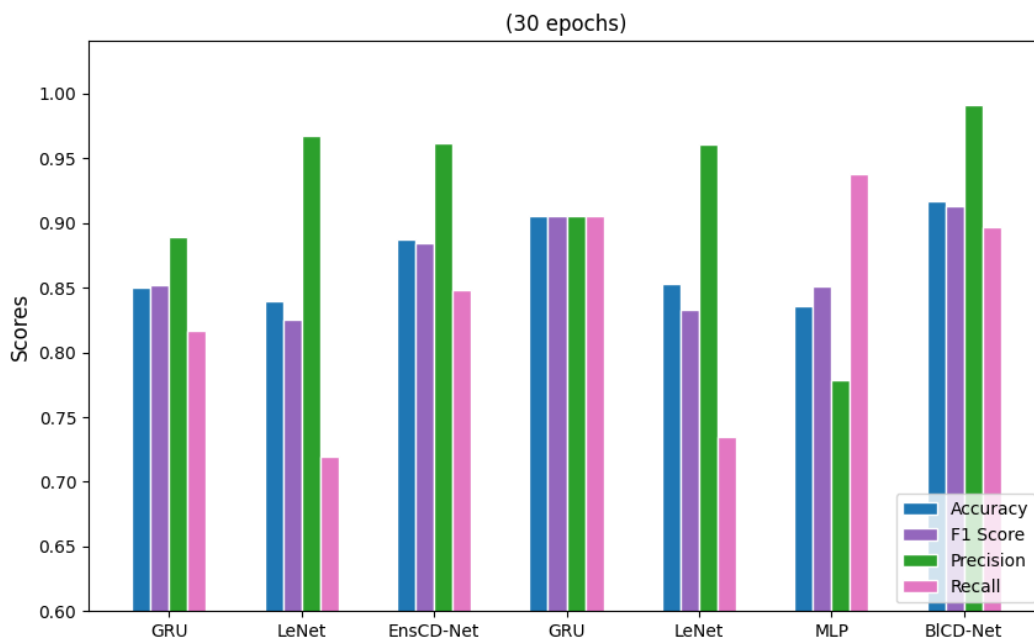
an F1-score of 90.7%. Table 13 shows that BICVDD-Net exhibits the highest performance, followed by LeNet and MLP. Execution times are well within the acceptable limits.

At 30 epochs, the performance of BICVDD-Net enhances, with an accuracy of 91.7% and an F1-score of 91.2%. Moreover, it consistently outperforms the other base models,

showcasing its effectiveness. Execution times, especially for the BICVDD-Net, remain reasonable considering the substantial performance improvement, which can be seen in Table 14. For enhancing the generalizability of the proposed DL model, 10-FCV is implemented. In all these tables, the BICVDD-Net has three values for each performance metric.



**FIGURE 9.** Simulation representation of ensemble and blending based cardiovascular disease detection networks on 20 epochs for heart disease prediction.



**FIGURE 10.** Simulation representation of ensemble and blending based cardiovascular disease detection networks on 30 epochs for heart disease prediction.

The first bold value is the mean value while the other values (enclosed in square brackets) show 95% CI measured from 10-FCV results.

Overall, BICVDD-Net, which is the essence of three different DL models (LeNet, GRU and MLP), proves to be a powerful blending approach for heart disease prediction.

It consistently demonstrates the best performance, indicating its ability to effectively capture and leverage the insights from the base models. The results also suggest that increasing the number of training epochs generally leads to improved performance across all models. However, it is essential to consider the trade-off between computational resources



**TABLE 4.** Hyper-parameters utilized in ensemble based cardiovascular disease detection network.

Common Parameters:	
Loss Function	Binary cross_entropy
Metrics	Accuracy
Epoch	5-30
Batch Size	32
LeNet:	
Optimizer	SGD
Number of neurons in each layer	4-64
Activation Function on Output Layer	Sigmoid
GRU:	
Optimizer	SGD
Number of neurons in each layer	16
Activation Function on Output Neuron	Sigmoid
EnsCVDD-Net:	
Optimizer	SGD
Number of neurons in each layer	4-64
Activation Function on Output Neuron	Sigmoid

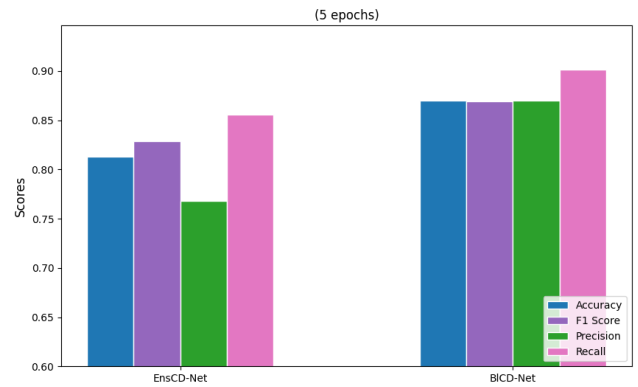
(as indicated by execution time) and the marginal gains in performance. The combined simulation results of EnsCVDD-Net and BICVDD-Net at different epochs are shown in Figs. 7, 8, 9 and 10.

**J. COMPARISON OF ENSEMBLE AND BLENDING BASED CARDIOVASCULAR DISEASE DETECTION NETWORKS**

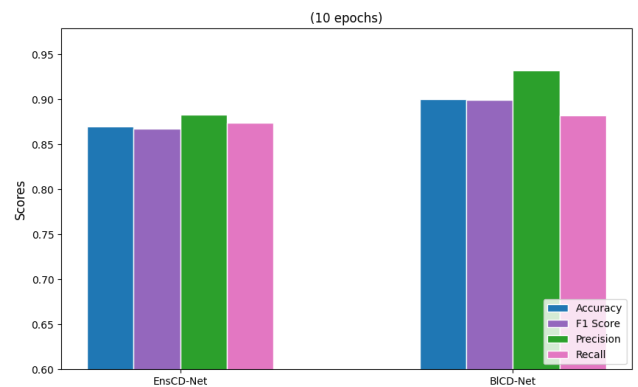
In the proposed system, two DL solutions, EnsCVDD-Net and BICVDD-Net, have been proposed for heart disease prediction. Both the networks are implemented on the same dataset. EnsCVDD-Net utilizes a parallel hybrid combination of LeNet and GRU. Both LeNet and GRU are trained in parallel, and their outputs are concatenated before the final prediction. While BICVDD-Net combines the predictions of two base models (LeNet and GRU) using a meta-model (MLP). It combines the outputs of the base models and feeds them to the meta-model for final prediction. The base models (LeNet and GRU) are trained separately on validation set, after which the prediction of these base models is concatenated and the meta-model (MLP) is trained on it.

The precision, recall, accuracy and F1-score metrics are used to gauge the two proposed models. A comparison of the evaluation results of EnsCVDD-Net and BICVDD-Net are shown in Figs. 11, 12, 13 and 14. With the blend of three models, BICVDD-Net performs better than EnsCVDD-Net with an accuracy of 87%, precision of 87.03%, recall of 86.9%, and F1-score of 86.9% at 5 epochs (Fig. 11). BICVDD-Net maintains its good performance at 10 epochs in all measures with an accuracy of 89.4%, when compared with the performance of EnsCVDD-Net, as shown in Fig. 12. Similarly, with a competitive accuracy of 91.1% and 91.7% in Figs. 13 and 14 shows the consistent performance of BICVDD-Net over EnsCVDD-Net at 20 and 30 epochs, respectively.

Based on the evaluation results, BICVDD-Net outperforms EnsCVDD-Net across all performance metrics. This suggests



**FIGURE 11.** Comparison of ensemble and blending based cardiovascular disease detection networks at 5 epochs.



**FIGURE 12.** Comparison of ensemble and blending based cardiovascular disease detection networks at 10 epochs.

that the approach of blending predictions from LeNet and GRU using an MLP meta-model yields better results as compared to the parallel combination of LeNet and GRU. It offers better generalization by capturing diverse patterns from both models and effectively combining their strengths. Additionally, the architecture of BICVDD-Net provides more flexibility in integrating different types of models, potentially leading to better performance in heart disease prediction.

**K. SHAPLEY ADDITIVE EXPLANATIONS**

SHAP is an interpretability framework that helps in understanding the model’s output. It works on the principle of game theory and it assigns a significance value to every feature of the model [59]. This helps in explaining the role each feature plays in the prediction process. SHAP provides a clear picture of the role each feature plays in the prediction process. In this way, it not only enhances the interpretability of ML models but also ensures fairness [60].

In terms of the BRFS dataset, the individual feature values are considered as the players. The SHAP values show how much each feature has contributed to predicting CVD.

In this study, a kernel explainer is used with SHAP. A kernel explainer uses local LR with specific weights to

**TABLE 5.** Simulation Results (with Confidence Interval of 95%) of ensemble based cardiovascular disease detection network (5 Epochs).

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Execution Time (sec)
GRU	76.0	73.8	82.9	78.1	284.10
LeNet	75.9	68.3	95.8	79.8	100.92
EnsCVDD-Net	<b>79.2</b> , [78.0, 81.0]	<b>77.0</b> , [73.0, 81.0]	<b>86.2</b> , [81.2, 91.1]	<b>81.0</b> , [80.0, 82.0]	<b>143.3</b> , [136.4, 150.2]

**TABLE 6.** Simulation results (with confidence interval of 95%) of ensemble based cardiovascular disease detection network (10 Epochs).

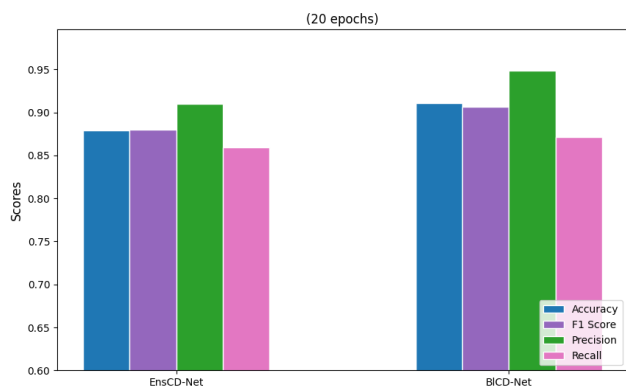
Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Execution Time (sec)
GRU	78.7	77.9	79.7	78.8	494.03
LeNet	81.1	97.7	63.5	77.0	206.68
EnsCVDD-Net	<b>85.1</b> , [84.0, 87.0]	<b>85.0</b> , [81.0, 89.0]	<b>87.2</b> , [84.0, 91.0]	<b>86.0</b> , [85.0, 86.3]	<b>295.2</b> , [277.0, 313.6]

**TABLE 7.** Simulation results (with confidence interval of 95%) of ensemble based cardiovascular disease detection network (20 Epochs).

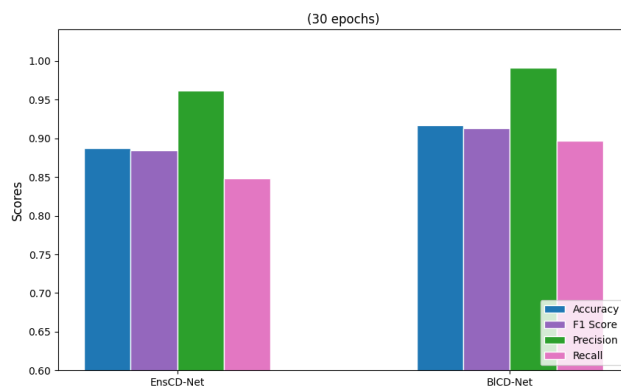
Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Execution Time (sec)
GRU	82.7	83.1	81.9	82.5	785.39
LeNet	85.0	83.4	91.0	87.0	326.03
EnsCVDD-Net	<b>87.0</b> , [86.3, 88.0]	<b>90.0</b> , [87.0, 93.0]	<b>85.0</b> , [81.3, 88.0]	<b>87.0</b> , [86.2, 88.0]	<b>505.1</b> , [495.0, 515.5]

**TABLE 8.** Simulation results (with confidence interval of 95%) of ensemble based cardiovascular disease detection network (30 Epochs).

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Execution Time (sec)
GRU	85.0	88.9	81.7	85.2	971.11
LeNet	84.0	96.7	71.9	82.5	507.96
EnsCVDD-Net	<b>88.0</b> , [87.2, 89.0]	<b>91.0</b> , [88.1, 94.0]	<b>85.0</b> , [82.10, 87.0]	<b>87.3</b> , [87.0, 88.0]	<b>760.3</b> , [746.0, 775.1]



**FIGURE 13.** Comparison of ensemble and blending based cardiovascular disease detection networks at 20 epochs.



**FIGURE 14.** Comparison of ensemble and blending based cardiovascular disease detection networks at 30 epochs.

compute the SHAP values. This kernel is applied to all the features of the dataset and it checks how much each feature is contributing to the predicted outcome. Once the SHAP values are calculated for each feature, various visualizations are generated. These graphs provide a better explanation of these contributions. In this way, SHAP helps in understanding which features are important and how they interact with each other. Figs. 15 and 20 show the summary plot for EnsCVDD-Net and BICVDD-Net, where y-axis includes the

list of features contributing to the prediction and x-axis shows the average SHAP value for each feature. The color of the bar indicates the distribution of individual SHAP values for each feature.

The beeswarm plot [61] provides a more detailed view of how individual features contribute to the model’s predictions than the average SHAP values in the bar plot. It is used to visualize the distribution of individual SHAP values for each feature and helps to understand how the model’s predictions

**TABLE 9. Architectures of deep learning models and blending based cardiovascular disease detection network.**

Models	Architecture
LeNet	Conv 1D layer(Number of neurons=32, kernel_size=5, activation='tanh') MaxPooling layer(padding = 'same') Conv 1D layer(Number of neurons=16, kernel_size=5, activation='tanh') MaxPooling(padding= 'same') Dense layer(Number of neuron=64, activation='tanh') Dense layer(Number of neurons=32, activation='tanh') Dense layer(num_classes, activation='sigmoid')
GRU	GRU(Number of neurons=16, activation='tanh', return_sequences=False) Dense layer(Number of neurons=1, activation function='sigmoid')
MLP	Dense layer(Number of neurons= 256, activation='tanh') Dense layer(Number of neurons= 128, activation='tanh') Dense layer(Number of neurons= 64, activation='tanh') Dense layer(Number of neuron= 32, activation='tanh') Dense layer(num_classes, activation='sigmoid')
BICVDD-Net	Conv 1D layer(Number of neurons=32, kernel_size=5, activation='tanh') MaxPooling layer(padding = 'same') Conv 1D layer(Number of neurons=16, kernel_size=5, activation='tanh') MaxPooling(padding= 'same') Dense layer(Number of neuron=64, activation='tanh') Dense layer(Number of neurons=32, activation='tanh') GRU(Number of neurons=16, activation='tanh', return_sequences=False) Dense layer(Number of neurons= 256, activation='tanh') Dense layer(Number of neurons= 128, activation='tanh') Dense layer(Number of neurons= 64, activation='tanh') Dense layer(Number of neuron= 32, activation='tanh') Dense layer(Number of neurons=2, activation function='sigmoid')

vary for different values of a particular feature. Here Figs. 17 and 22 are the beeswarm plots for EnsCVDD-Net and BICVDD-Net, respectively.

ML models behave as a black box. To understand the output of the model and to interpret the influence of input features, a waterfall plot is used in ML [62]. In this scenario, it is used to understand the impact of various health factors on the outcomes of EnsCVDD-Net and BICVDD-Net for predicting CVD (Fig. 16 and Fig. 21).

A SHAP dependence plot is a type of plot that shows how a feature's value affects a model's prediction. It shows the average relationship between a feature and the model's prediction [63]. However, it does not show how the relationship varies for individual data points. In Fig. 18, the feature is 'GenHlth' and the EnsCVDD-Net model's prediction is a

**TABLE 10. Hyper-parameters utilized in blending based cardiovascular disease detection network.**

Common Parameters:	
Loss Function	Binary cross_entropy
Metrics	Accuracy
Epoch	5-30
Batch Size	32
<b>LeNet:</b>	
Optimizer	Adam
Number of neurons in each layer	16-64
Activation Function on Output Layer	Sigmoid
<b>GRU:</b>	
Optimizer	Adam
Number of neurons in each layer	16
Activation Function on Output Neuron	Sigmoid
<b>MLP:</b>	
Optimizer	Adam
Number of neurons in each layer	32-256
Activation Function on Output Neuron	Sigmoid
<b>BICVDD-Net:</b>	
Optimizer	Adam
Number of neurons in each layer	16-256
Activation Function on Output Neuron	Sigmoid

SHAP value, which is a measure of how much a particular feature contributes to the model's prediction. The color of the dots represents the density of the data points, with darker colors representing more data points. The plot shows that there is a positive relationship between the feature 'GenHlth' and the SHAP value. This means that as the value of 'GenHlth' increases, the SHAP value also increases. In other words, the 'GenHlth' feature has a positive impact on the model's prediction. Similarly, Fig. 23 show dependence plot for BICVDD-Net.

The force plot is a type of Shapley value plot, which is used to understand the contribution of each feature to the prediction of a specific instance in an ML model. In the force plot, the base value is considered to be the average effect of all the features. Each feature either pushes the prediction higher (red) or lower (blue) from this base value. The final prediction (f(x)) is the sum of the base value and all the feature contributions. Fig. 19 and Fig. 24 represent the force plots for EnsCVDD-Net and BICVDD-Net, respectively. In Fig. 24, the feature 'Age' has the largest positive impact on the heart disease prediction, while the feature 'BMI' has the largest negative impact. However, the positive contributions of features 'HighChol', 'Sex', and 'Age' are not enough to offset the negative contributions of 'Income', 'PhysHlth', and 'BMI', resulting in a final prediction of less than the base value.

**L. COMPARISON OF SHAPLEY ADDITIVE EXPLANATIONS PLOTS**

To learn the feature contribution in the decision making process of the proposed system model, SHAP is implemented on EnsCVDD-Net and BICVDD-Net. From the SHAP plots of EnsCVDD-Net for predicting heart disease, it is observed

**TABLE 11.** Simulation results (with confidence interval of 95%) of blending based cardiovascular disease detection network (5 Epochs).

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Execution Time (sec)
GRU	82.5	90.5	72.6	80.6	84.27
LeNet	82.6	80.8	85.5	83.1	23.25
MLP	74.9	68.7	91.4	78.5	43.26
BICVDD-Net	<b>85.0</b> , [84.0, 85.2]	<b>84.00</b> , [82.40, 85.60]	<b>86.0</b> , [84.0, 88.0]	<b>85.0</b> , [84.0, 85.1]	<b>37.1</b> , [34.0, 41.0]

**TABLE 12.** Simulation results (with confidence interval of 95%) of blending based cardiovascular disease detection network (10 Epochs).

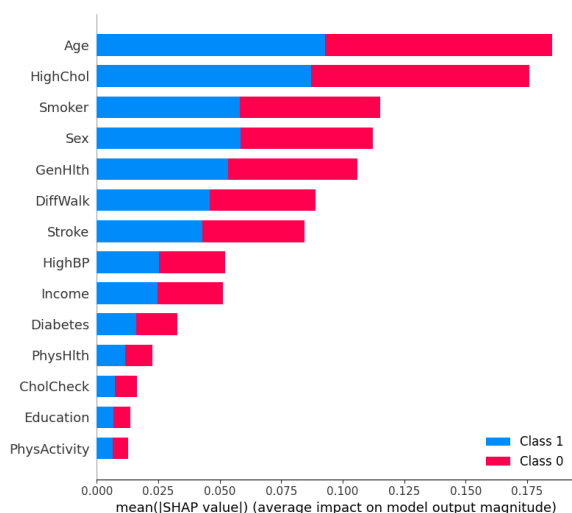
Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Execution Time (sec)
GRU	86.0	98.8	72.9	83.9	144.10
LeNet	85.1	84.5	85.9	85.2	84.55
MLP	78.0	72.9	89.1	80.2	84.29
BICVDD-Net	<b>89.7</b> , [89.5, 89.9]	<b>93.0</b> , [91.7, 94.2]	<b>86.0</b> , [85.0, 87.0]	<b>89.4</b> , [89.2, 89.6]	<b>73.2</b> , [64.8, 81.5]

**TABLE 13.** Simulation results (with confidence interval of 95%) of blending based cardiovascular disease detection network (20 Epochs).

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Execution Time (sec)
GRU	89.0	97.7	80.8	88.5	264.26
LeNet	88.0	89.9	86.9	88.3	84.06
MLP	82.7	87.4	76.5	81.6	77.07
BICVDD-Net	<b>90.7</b> , [90.4, 91.0]	<b>95.0</b> , [94.0, 97.0]	<b>86.0</b> , [85.0, 87.3]	<b>90.2</b> , [89.9, 90.5]	<b>125.0</b> , [103.3, 146.3]

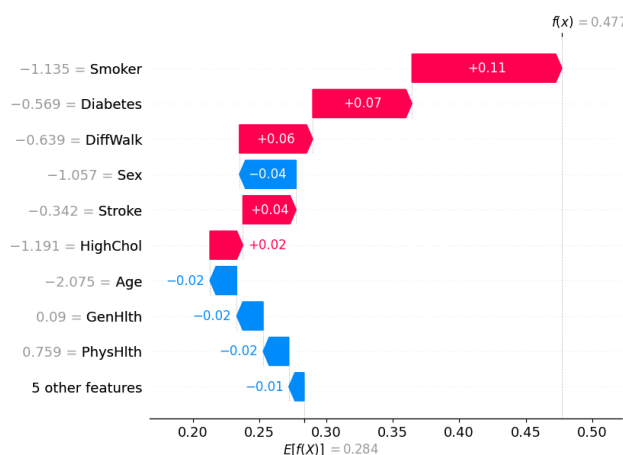
**TABLE 14.** Simulation results (with confidence interval of 95%) of blending based cardiovascular disease detection network (30 Epochs).

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Execution Time (sec)
GRU	90.5	90.5	90.5	90.5	345.26
LeNet	85.3	96.0	73.5	83.3	144.8
MLP	83.6	77.9	93.8	85.1	144.21
BICVDD-Net	<b>91.5</b> , [91.0, 92.0]	<b>96.0</b> , [94.0, 97.3]	<b>87.0</b> , [85.0, 89.0]	<b>90.7</b> , [90.1, 91.2]	<b>236.5</b> , [204.6, 268.4]



**FIGURE 15.** Summary Plot of ensemble based cardiovascular disease detection network.

that the features like ‘Age’ and ‘High Cholesterol’ have the highest impact in generating heart disease prediction. ‘Sex’, ‘GenHlth’, ‘DiffWalk’, ‘Stroke’ and ‘Smoker’ have



**FIGURE 16.** Waterfall plot of ensemble based cardiovascular disease detection network.

considerable impacts on the model’s predictions. While features like ‘HighBP’, ‘Income’, ‘Diabetes’, and ‘PhysHlth’ have low impact, with the least contributing features to be ‘CholCheck’, ‘Education’ and ‘PhysActivity’ for EnsCVDD-Net in forecasting heart disease. It means higher

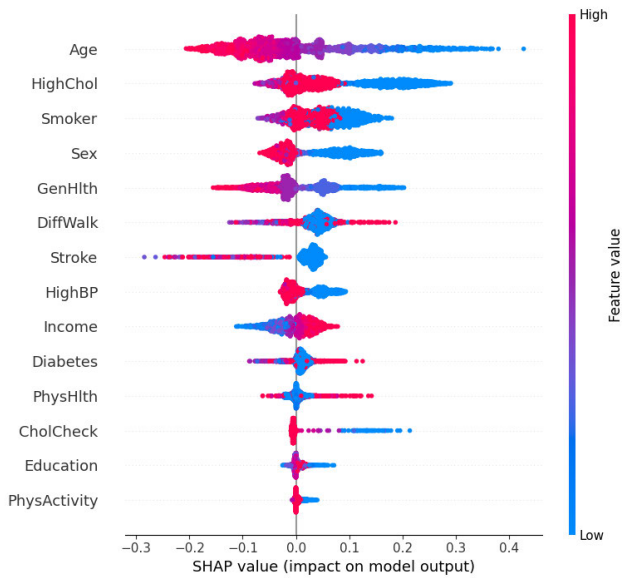


FIGURE 17. Beeswarm plot of ensemble based cardiovascular disease detection network.

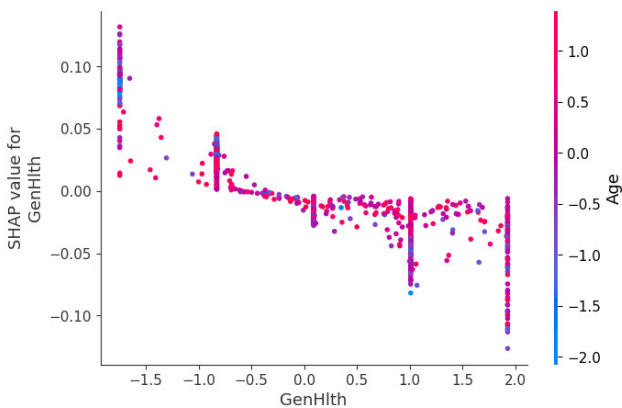


FIGURE 18. Dependence plot of ensemble based cardiovascular disease detection network.

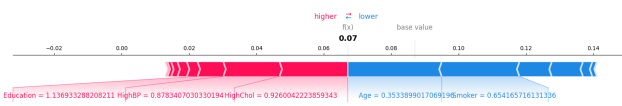


FIGURE 19. Force plot of ensemble based cardiovascular disease detection network.

values of age (old age individuals) have a higher chance of getting the heart disease. Similarly, individuals with high cholesterol, heavy smoking habits, poor general health, difficulty in walking or having a stroke significantly increase the chance of heart disease. It also coincides with the general interpretation for predicting heart disease, i.e., there is a high chance of getting heart disease with the increase in age factor, high BP, high cholesterol levels, poor general health, or heart stroke, etc.

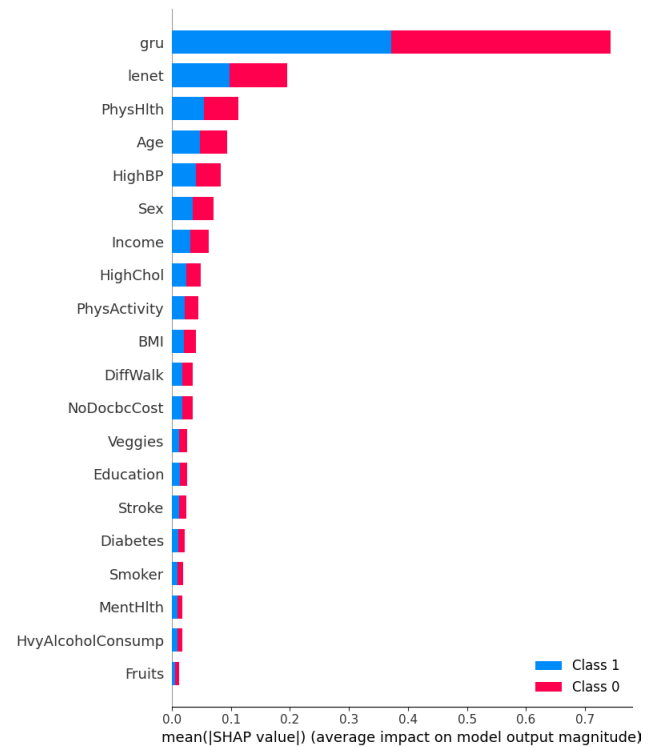


FIGURE 20. Summary plot of blending based cardiovascular disease detection network.

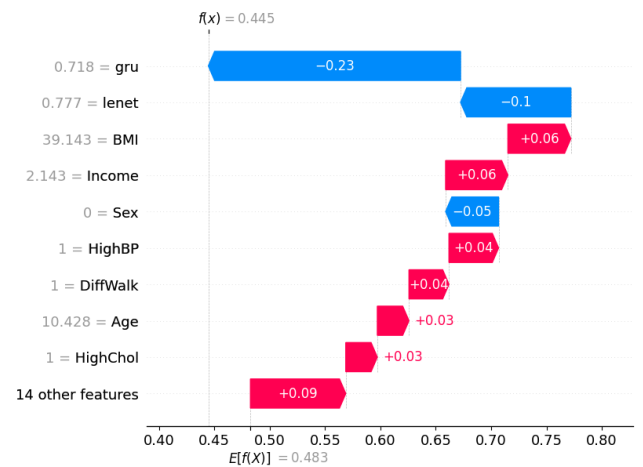


FIGURE 21. Waterfall plot of blending based cardiovascular disease detection network.

In a similar manner, the SHAP plots of BICVDD-Net for heart disease prediction shows the predictions of the base models. GRU and LeNet play significant contribution for BICVDD-Net in early prediction of CVD. While health related features like ‘PhysHlth’, ‘Age’, ‘HighBP’, ‘Sex’ and ‘Income’ play a critical role in the decision making process for BICVDD-Net. Other features including ‘Veggies’, ‘Education’, ‘Stroke’, ‘Diabetes’, ‘Smoker’, ‘MentHlth’, ‘HvyAlcoholConsump’ and ‘Fruits’ have low contributions

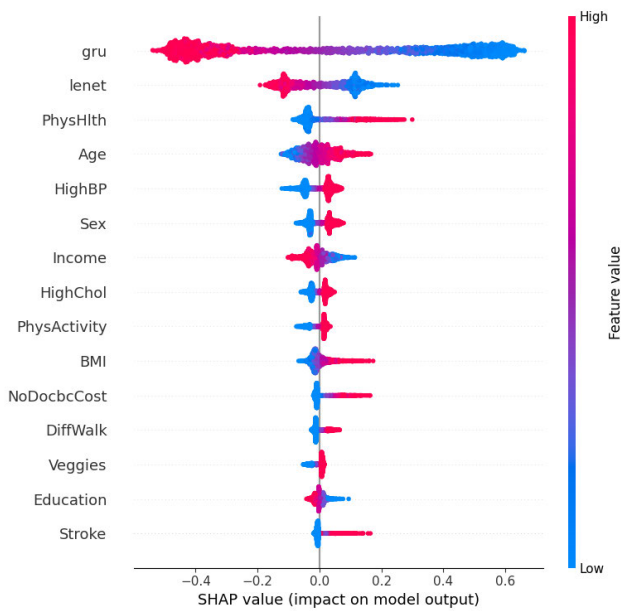


FIGURE 22. Beeswarm plot of blending based cardiovascular disease detection network.

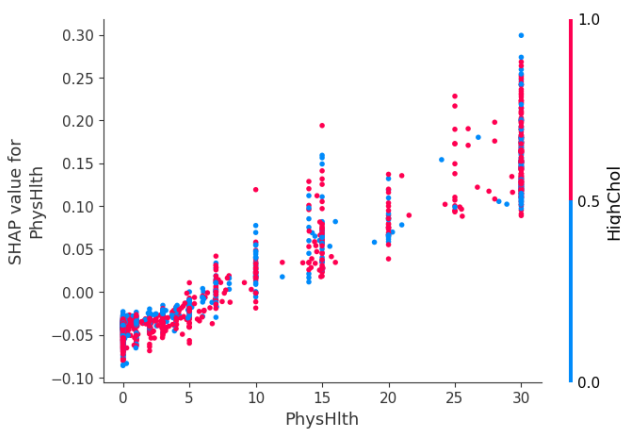


FIGURE 23. Dependence plot of blending based cardiovascular disease detection network.

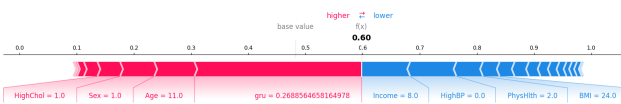


FIGURE 24. Force plot of blending based cardiovascular disease detection network.

to the proposed BICVDD-Net model’s output. Comparing the SHAP plots of both the proposed models show that ‘Age’, ‘Sex’, ‘HighChol’, ‘HighBP’, ‘PhyHlth’, and ‘Stroke’ are important features for predicting CVD.

**VI. CONCLUSION**

The aim this study is to accurately predict CVDs while minimizing the difficulties caused by unbalanced data and

taking the patient’s health and socioeconomic circumstances into account. This study utilizes the Heart Disease Health Indicator dataset. The ADASYN data balancing approach is employed due to the high imbalance ratio in this dataset. Furthermore, feature selection is done using PBCC. In this article, EnsCVDD-Net and BICVDD-Net are constructed for heart disease prediction. EnsCVDD-Net is a hybrid combination of LeNet and GRU. BICVDD-Net incorporates LeNet and GRU as base-models and MLP as a meta-model. The findings indicate that the EnsCVDD-Net achieves 88% accuracy, 88% F1-score, 91% precision, 85% recall, and 777s execution time, surpassing all base models. Similarly, BICVDD-Net outperforms its base models with 91% accuracy, 91% F1-score, 96% precision, 86% recall, and 247s execution time. 10-FCV is also used to estimate model prediction and adjust model parameters. To see each feature’s contribution to the output, SHAP is utilized. SHAP values break down the output and summarize each feature’s behavior in the network prediction. In the future, this study can be implemented on multiple healthcare datasets to the proposed system models generalizable. Additionally, by implementing the proposed system models in a real-time hospital setting, continuous monitoring of heart disease patient data will allow the proposed system models to evolve over time. This dynamic adaptation will enhance the accuracy and responsiveness of the EnsCVDD-Net and BICVDD-Net, resulting in improved patient outcomes.

**APPENDIX  
LIST OF ABBREVIATIONS**

TABLE 15. List of abbreviations.

Abbreviation	Description
10-FCV	10-Fold Cross Validation
ADASYN	Adaptive Synthetic Sampling Technique
AI	Artificial Intelligence
BICVDD-Net	Blending based Cardiovascular Disease Detection Network
BP	Blood Pressure
BRFSS	Behavioural Risk Factor Surveillance System
CDC	Centers for Disease Control
CHD	Chronic Heart Disease
CNN	Convolutional Neural Network
CVD	Cardiovascular Disease
DL	Deep Learning
DNN	Deep Neural Network
EnsCVDD-Net	Ensemble based Cardiovascular Disease Detection Network
GRU	Gated Recurrent Unit
LSTM	Long Short Term Memory
ML	Machine Learning
MLP	Multilayer Perceptron
PBCC	Point Biserial Correlation Coefficient
RNN	Recurrent Neural Network
ROS	Random Over Sampling
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Oversampling Technique
XAI	eXplainable Artificial Intelligence

## ACKNOWLEDGMENT

The authors extend their appreciation to the Researchers Supporting Program for Project number (RSPD2024R648), at King Saud University for supporting this research project.

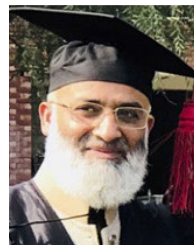
## REFERENCES

- [1] M. B. Abubaker and B. Babayigit, "Detection of cardiovascular diseases in ECG images using machine learning and deep learning methods," *IEEE Trans. Artif. Intell.*, vol. 4, no. 2, pp. 373–382, Apr. 2023.
- [2] S. Rehman, E. Rehman, M. Ikram, and Z. Jianglin, "Cardiovascular disease (CVD): Assessment, prediction and policy implications," *BMC Public Health*, vol. 21, no. 1, pp. 1–14, Dec. 2021.
- [3] M. Inam, Z. Samad, E. M. Vaughan, A. Almas, B. Hanif, A. M. Minhas, Z. Jarrar, F. Z. Habib, S. Sheikh, D. Zhu, and S. S. Virani, "Global cardiovascular research: Gaps and opportunities," *Current Cardiology Rep.*, vol. 25, no. 12, pp. 1831–1838, Dec. 2023.
- [4] F. Agyekum, A. A. Folsom, B. Abaidoo, L. T. Appiah, Y. Adu-Boakye, H. Ayetey, and I. K. Owusu, "Behavioural and nutritional risk factors for cardiovascular diseases among the Ghanaian population—A cross-sectional study," *BMC Public Health*, vol. 24, no. 1, p. 194, Jan. 2024.
- [5] A. Kumar, K. U. Singh, and M. Kumar, "A clinical data analysis based diagnostic systems for heart disease prediction using ensemble method," *Big Data Mining Analytics*, vol. 6, no. 4, pp. 513–525, Dec. 2023.
- [6] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: A systematic literature review, synthesizing framework and future research agenda," *J. Ambient Intell. Humanized Comput.*, vol. 14, no. 7, pp. 8459–8486, Jul. 2023.
- [7] A. P. Jawalkar, P. Swetcha, N. Manasvi, P. Sreekalaa, S. Aishwarya, P. Kanaka Durga Bhavani, and P. Anjani, "Early prediction of heart disease with data analysis using supervised learning with stochastic gradient boosting," *J. Eng. Appl. Sci.*, vol. 70, no. 1, p. 122, Dec. 2023.
- [8] S. Nazir, S. Khan, H. U. Khan, S. Ali, I. García-Magariño, R. B. Atan, and M. Nawaz, "A comprehensive analysis of healthcare big data management, analytics and scientific programming," *IEEE Access*, vol. 8, pp. 95714–95733, 2020.
- [9] Y. Shen, J. Zhu, Z. Deng, W. Lu, and H. Wang, "EnsDeepDP: An ensemble deep learning approach for disease prediction through metagenomics," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 20, no. 2, pp. 986–998, Mar. 2023.
- [10] A. Nicolucci, L. Romeo, M. Bernardini, M. Vespasiani, M. C. Rossi, M. Petrelli, A. Ceriello, P. Di Bartolo, E. Frontoni, and G. Vespasiani, "Prediction of complications of type 2 diabetes: A machine learning approach," *Diabetes Res. Clin. Pract.*, vol. 190, Aug. 2022, Art. no. 110013.
- [11] V. Chang, M. A. Ganatra, K. Hall, L. Golightly, and Q. A. Xu, "An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators," *Healthcare Analytics*, vol. 2, Nov. 2022, Art. no. 100118.
- [12] R. Rajalakshmi, P. Sivakumar, L. K. Kumari, and M. C. Selvi, "A novel deep learning model for diabetes mellitus prediction in IoT-based healthcare environment with effective feature selection mechanism," *J. Supercomput.*, vol. 80, no. 1, pp. 271–291, Jan. 2024.
- [13] N. El-Rashidy, N. E. ElSayed, A. El-Ghamry, and F. M. Talaat, "Utilizing fog computing and explainable deep learning techniques for gestational diabetes prediction," *Neural Comput. Appl.*, vol. 35, no. 10, pp. 7423–7442, Apr. 2023.
- [14] T. Asuroglu and H. Ogul, "A deep learning approach for Parkinson's disease severity assessment," *Health Technol.*, vol. 12, no. 5, pp. 943–953, 2022.
- [15] M. A. H. Saeed, "Diabetes type 2 classification using machine learning algorithms with up-sampling technique," *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 1, pp. 1–10, Feb. 2023.
- [16] C. J. Ejyiyi, Z. Qin, J. Amos, M. B. Ejyiyi, A. Nnani, T. U. Ejyiyi, V. K. Agbesi, C. Diokpo, and C. Okpara, "A robust predictive diagnosis model for diabetes mellitus using Shapley-incorporated machine learning algorithms," *Healthcare Analytics*, vol. 3, Nov. 2023, Art. no. 100166.
- [17] J. C. T. Arroyo and A. J. P. Delima, "An optimized neural network using genetic algorithm for cardiovascular disease prediction," *J. Adv. Inf. Technol.*, vol. 13, no. 1, 2022.
- [18] D. Albers, M. Sirlanci, M. Levine, J. Claassen, C. Der Nigoghossian, and G. Hripscak, "Interpretable physiological forecasting in the ICU using constrained data assimilation and electronic health record data," *J. Biomed. Inform.*, vol. 145, Sep. 2023, Art. no. 104477.
- [19] V. Chaurasia and A. Chaurasia, "Novel method of characterization of heart disease prediction using sequential feature selection-based ensemble technique," *Biomed. Mater. Devices*, vol. 1, no. 2, pp. 932–941, Sep. 2023.
- [20] E. Longato, G. P. Fadini, G. Sparacino, A. Avogaro, L. Tramontan, and B. Di Camillo, "A deep learning approach to predict diabetes' cardiovascular complications from administrative claims," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 9, pp. 3608–3617, Sep. 2021.
- [21] K. Hymavathi, S. H. Mehanoor, G. Srinivas, A. Lakshmanarao, and J. Jalla, "Advancing heart disease detection by using ensemble meta-features integration," in *Proc. IEEE Int. Conf. Comput., Power Commun. Technol. (IC2PCT)*, Feb. 2024, pp. 888–892.
- [22] P. Bizopoulos and D. Koutsouris, "Deep learning in cardiology," *IEEE Rev. Biomed. Eng.*, vol. 12, pp. 168–193, 2019.
- [23] T. M. A. Monisha Sharean and G. Johncy, "Deep learning models on heart disease Estimation—A review," *J. Artif. Intell. Capsule Netw.*, vol. 4, no. 2, pp. 122–130, Jul. 2022.
- [24] Y. Zheng and X. Hu, "Healthcare predictive analytics for disease progression: A longitudinal data fusion approach," *J. Intell. Inf. Syst.*, vol. 55, no. 2, pp. 351–369, Oct. 2020.
- [25] M. Alfreihat, O. S. Almousa, Y. Tashtoush, A. AlSobeh, K. Mansour, and H. Migdady, "Emo-SL framework: Emoji sentiment lexicon using text-based features and machine learning for sentiment analysis," *IEEE Access*, vol. 12, pp. 81793–81812, 2024.
- [26] T. Phan, D. Le, P. Brijesh, D. Adjeroh, J. Wu, M. O. Jensen, and N. Le, "Multimodality multi-lead ECG arrhythmia classification using self-supervised learning," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Sep. 2022, pp. 1–4.
- [27] S. Alshattawi, A. Shatnawi, A. M. R. AlSobeh, and A. A. Magableh, "Beyond word-based model embeddings: Contextualized representations for enhanced social media spam detection," *Appl. Sci.*, vol. 14, no. 6, p. 2254, Mar. 2024.
- [28] L. Rasmy, Y. Xiang, Z. Xie, C. Tao, and D. Zhi, "Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction," *NPJ Digit. Med.*, vol. 4, no. 1, p. 86, May 2021.
- [29] X. Xu, T. Wang, J. Zhuang, H. Yuan, M. Huang, J. Cen, and Y. Shi, "ImageCHD: A 3D computed tomography image dataset for classification of congenital heart disease," in *Proc. 23rd Int. Conf., Lima, Peru, Cham, Switzerland: Springer 2020*, pp. 77–87.
- [30] J. Liao, D. Liu, G. Su, and L. Liu, "Recognizing diseases with multivariate physiological signals by a DeepCNN-LSTM network," *Appl. Intell.*, vol. 51, no. 11, pp. 7933–7945, Nov. 2021.
- [31] Y. Bouchareb, P. Moradi Khaniabadi, F. Al Kindi, H. Al Dhuhli, I. Shiri, H. Zaidi, and A. Rahmim, "Artificial intelligence-driven assessment of radiological images for COVID-19," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104665.
- [32] P. Dileep, K. N. Rao, P. Bodapati, S. Gokuruboyina, R. Peddi, A. Grover, and A. Sheetal, "An automatic heart disease prediction using cluster-based bi-directional LSTM (C-BiLSTM) algorithm," *Neural Comput. Appl.*, vol. 35, no. 10, pp. 7253–7266, Apr. 2023.
- [33] V. K. Sudha and D. Kumar, "Hybrid CNN and LSTM network for heart disease prediction," *Social Netw. Comput. Sci.*, vol. 4, no. 2, p. 172, Jan. 2023.
- [34] N. Sharma, L. Malviya, A. Jadhav, and P. Lalwani, "A hybrid deep neural net learning model for predicting coronary heart disease using randomized search cross-validation optimization," *Decis. Analytics J.*, vol. 9, Dec. 2023, Art. no. 100331.
- [35] J. Naskath, G. Sivakamasundari, and A. A. S. Begum, "A study on different deep learning algorithms used in deep neural nets: MLP SOM and DBN," *Wireless Pers. Commun.*, vol. 128, no. 4, pp. 2913–2936, Feb. 2023.
- [36] E. A. Ogundepo and W. B. Yahya, "Performance analysis of supervised classification models on heart disease prediction," *Innov. Syst. Softw. Eng.*, vol. 19, no. 1, pp. 129–144, Mar. 2023.
- [37] A. Radwan, M. Amarnah, H. Alawneh, H. I. Ashqar, A. AlSobeh, and A. A. R. Magableh, "Predictive analytics in mental health leveraging LLM embeddings and machine learning models for social media analysis," *Int. J. Web Services Res.*, vol. 21, no. 1, pp. 1–22, Feb. 2024.
- [38] S. Praveenkumar, V. Veeraiah, S. Pramanik, S. M. Basha, A. V. L. Neto, V. H. C. De Albuquerque, and A. Gupta, "Prediction of patients' incurable diseases utilizing deep learning approach," in *Proc. Int. Conf. Innov. Comput. Commun.* Singapore: Springer Nature, 2023, pp. 33–44.

- [39] B. S. Shukur and M. M. Mijwil, "Involving machine learning techniques in heart disease diagnosis: A performance analysis," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 13, no. 2, p. 2177, Apr. 2023.
- [40] R. Kapila, T. Ragunathan, S. Saleti, T. J. Lakshmi, and M. W. Ahmad, "Heart disease prediction using novel quine McCluskey binary classifier (QMBC)," *IEEE Access*, vol. 11, pp. 64324–64347, 2023.
- [41] S. Dhanka, V. K. Bhardwaj, and S. Maini, "Comprehensive analysis of supervised algorithms for coronary artery heart disease detection," *Expert Syst.*, vol. 40, no. 7, p. e1330, Aug. 2023.
- [42] E. Kim, J. Kim, J. Park, H. Ko, and Y. Kyung, "TinyML-based classification in an ECG monitoring embedded system," *Comput., Mater. Continua*, vol. 75, no. 1, pp. 1751–1764, 2023.
- [43] J. Wang, C. Rao, M. Goh, and X. Xiao, "Risk assessment of coronary heart disease based on cloud-random forest," *Artif. Intell. Rev.*, vol. 56, no. 1, pp. 203–232, Jan. 2023.
- [44] A. Teboul. (2022). *Heart Disease Health Indicators Dataset*. Kaggle. Accessed: Jul. 13, 2024. [Online]. Available: <https://www.kaggle.com/datasets/alexteboul/heart-disease-health-indicators-dataset>
- [45] GeeksforGeeks. (2022). *ML: Matrix plots in Seaborn*. GeeksforGeeks. Accessed: Jul. 13, 2024. [Online]. Available: <https://www.geeksforgeeks.org/ml-matrix-plots-in-seaborn/>
- [46] P. Mooijman, C. Catal, B. Tekinerdogan, A. Lommen, and M. Blokland, "The effects of data balancing approaches: A case study," *Appl. Soft Comput.*, vol. 132, Jan. 2023, Art. no. 109853.
- [47] R. Nian. (2019). *An Introduction to ADASYN (With Code!)*. Medium. Accessed: Jul. 13, 2024. [Online]. Available: <https://medium.com/@ruinian/an-introduction-to-adasyn-with-code-1383a5ece7aa>
- [48] G. Shukla, S. Singh, C. Dhule, R. Agrawal, S. Saraswat, A. Al-Rasheed, M. S. Alqahtani, and B. O. Soufiene, "Point biserial correlation symbiotic organism search nanoengineering based drug delivery for tumor diagnosis," *Sci. Rep.*, vol. 14, no. 1, p. 6530, Mar. 2024.
- [49] M. H. R. Bhatti, N. Javaid, B. Mansoor, N. Alrajeh, M. Aslam, and M. Asad, "New hybrid deep learning models to predict cost from healthcare providers in smart hospitals," *IEEE Access*, vol. 11, pp. 136988–137010, 2023.
- [50] *Collaborative Markdown Knowledge Base (No Date) HackMD*. Accessed: Jul. 13, 2024. [Online]. Available: <https://hackmd.io/>
- [51] I. Shaheen, N. Javaid, N. Alrajeh, Y. Asim, and S. Aslam, "Hi-le and HiTCL: Ensemble learning approaches for early diabetes detection using deep learning and explainable artificial intelligence," *IEEE Access*, vol. 12, pp. 66516–66538, 2024.
- [52] M. Zulqarnain, R. Ghazali, Y. M. M. Hassim, and M. Aamir, "An enhanced gated recurrent unit with auto-encoder for solving text classification problems," *Arabian J. Sci. Eng.*, vol. 46, no. 9, pp. 8953–8967, Sep. 2021.
- [53] M. Jaffar, S. Shafiq, N. Shahzadi, N. Alrajeh, M. Jamil, and N. Javaid, "Efficient deep learning models for predicting super-utilizers in smart hospitals," *IEEE Access*, vol. 11, pp. 87676–87693, 2023.
- [54] C. Bento. (2021). *Multilayer Perceptron Explained With a Real-Life Example and Python Code: Sentiment Analysis*. Medium. Accessed: Jul. 13, 2024. [Online]. Available: <https://towardsdatascience.com/multi-layer-perceptron-explained-with-a-real-life-example-and-python-code-sentiment-analysis-cb408ee93141>
- [55] M. Banoula. (2023). *An Overview on Multilayer Perceptron (MLP)*. Simplilearn.com. Accessed: Jul. 13, 2024. [Online]. Available: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/multilayer-perceptron>
- [56] N. Nayyer, N. Javaid, M. Akbar, A. Aldeghisheem, N. Alrajeh, and M. Jamil, "A new framework for fraud detection in Bitcoin transactions through ensemble stacking model in smart cities," *IEEE Access*, vol. 11, pp. 90916–90938, 2023.
- [57] A. M. P. Chacón, I. S. Ramírez, and F. P. G. Márquez, "K-nearest neighbour and K-fold cross-validation used in wind turbines for false alarm detection," *Sustain. Futures*, vol. 6, Dec. 2023, Art. no. 100132.
- [58] T. R. Mahesh, O. Geman, M. Margala, and M. Guduri, "The stratified K-folds cross-validation and class-balancing methods with high-performance ensemble classifiers for breast cancer classification," *Healthcare Analytics*, vol. 4, Dec. 2023, Art. no. 100247.
- [59] A. Noor, N. Javaid, N. Alrajeh, B. Mansoor, A. Khaqan, and S. H. Bouk, "Heart disease prediction using stacking model with balancing techniques and dimensionality reduction," *IEEE Access*, vol. 11, pp. 116026–116045, 2023.
- [60] Towards Data Science. *Toward Explainable AI (XAI) With SHAP for Multi-Class Classification Problem*. Accessed: Mar. 13, 2024. [Online]. Available: <https://towardsdatascience.com/explainable-ai-xai-with-shap-multi-class-classification-problem>
- [61] *SHAP Documentation: Beeswarm Plot*. Accessed: Mar. 13, 2024. [Online]. Available: [https://shap.readthedocs.io/en/latest/example\\_notebooks/api\\_examples/plots/beeswarm.html](https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/beeswarm.html)
- [62] *SHAP Documentation: Waterfall Plot*. Accessed: Mar. 13, 2024. [Online]. Available: [https://shap.readthedocs.io/en/latest/example\\_notebooks/api\\_examples/plots/waterfall.html](https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/waterfall.html)
- [63] *SHAP-LRJBall Documentation: Dependence Plot*. Accessed: Mar. 13, 2024. [Online]. Available: [https://shap-lrjball.readthedocs.io/en/latest/example\\_notebooks/plots/dependence\\_plot.html](https://shap-lrjball.readthedocs.io/en/latest/example_notebooks/plots/dependence_plot.html)



**HIRA KHAN** received the bachelor's degree in electronics and the master's degree in electronic system engineering from COMSATS University Islamabad (CUI), Islamabad Campus, Pakistan, in 2011 and 2015, respectively, where she is currently pursuing the Ph.D. degree in deep learning and its applications in the healthcare sector using AI. She was a Research Associate with the Communications Over Sensors (ComSens) Research Laboratory, Department of Computer Science, CUI. Her research interests include deep learning, healthcare, and artificial intelligence.



**NADEEM JAVAID** received the Ph.D. degree from the University of Paris-Est, France, in 2010. He was a Visiting Professor with the University of Technology Sydney (UTS), Australia. He is currently a Tenured Professor and the Founding Director of the Communications Over Sensors (ComSens) Research Laboratory, Department of Computer Science, COMSATS University Islamabad, Islamabad Campus. He is a Visiting Professor with the International Graduate School of Artificial Intelligence, National Yunlin University of Science and Technology, Taiwan. He has supervised 203 master's and 34 Ph.D. theses. He has authored over 950 articles in technical journals and international conferences and U.S. Patent. His research interests include energy optimization in smart/microgrids, health care, and wireless sensor networks using blockchain and data analytics/AI. He received the Best University Teacher Award (BUTA'16) from the Higher Education Commission (HEC), Pakistan, in 2016, and the Research Productivity Award (RPA'17) from Pakistan Council for Science and Technology (PCST), in 2017. He has delivered many keynote and invited speeches at national and international conferences. With 28,000 citations and H-Index of 83, he is Pakistan's Best Scientist in Engineering and Technology/Computer Science in Stanford University's List of Top 2% Scientists in the World.



**TARIQ BASHIR** received the bachelor's degree in electrical and electronics engineering from the NWFP University of Engineering and Technology, Peshawar, Pakistan, in 2003, the master's degree in computer engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2009, and the Ph.D. degree in electrical engineering from COMSATS University Islamabad, Islamabad, Pakistan. He is currently an Associate Professor with COMSATS University Islamabad.

His research interests include digital watermarking and digital image processing.





**MARIAM AKBAR** received the M.Sc. and M.Phil. degrees from Quid-i-Azam University, Islamabad, and the Ph.D. degree in electrical engineering from the Communications Over Sensors (ComSens) Research Laboratory, COMSATS Univeristy Islamabad (CUI), Islamabad, in 2016, under the supervision of Prof. Nadeem Javaid. She has authored over 55 papers in technical journals and international conferences. She is currently a Tenured Associate Professor with the Department of Computer Science, CUI. Her research interests include wireless networks, smart grids, blockchain, data science, and AI.



**NABIL ALRAJEH** received the Ph.D. degree in biomedical informatics engineering from Vanderbilt University, USA. He was a Senior Advisor with the Ministry of Higher Education, his role was implementing development programs including educational affairs, strategic planning, and research and innovation. He was a member of the boards of trustees for five private universities in Saudi Arabia. He is currently a Professor of health informatics with King Saud University. His research interests include E-health applications, hospital information systems, telemedicine, healthcare applications of smart cities, and wireless sensor networks.



**SHERAZ ASLAM** received the B.S. degree in computer science from Bahauddin Zakariya University (BZU), Multan, Pakistan, in 2015, the M.S. degree in computer science with a specialization in energy optimization in the smart grid from COMSATS University Islamabad (CUI), Islamabad, Pakistan, in 2018, and the Ph.D. degree in computer engineering and informatics from Cyprus University of Technology (CUT), Limassol, Cyprus, under the supervision of Dr. Herodotos Herodotou. He was a Research Associate with Dr. Nadeem Javaid during the M.S. degree with CUI. He is currently a Postdoctoral Researcher with the DICL Research Laboratory, CUT, where he is also a part of an European Union funded research project named as aEROS. Previously, he was involved in several EU funded research projects, i.e., STEAM and MARI-Sense. He has authored more than 80 research publications in ISI-indexed international journals and conferences, including the IEEE INTERNET OF THINGS JOURNAL, *Renewable and Sustainable Energy Reviews*, and *Electric Power Systems Research*. His research interests include data analytics, generative adversarial networks, wireless networks, smart grids, and cloud computing. He is constantly looking for collaboration opportunities with professors and students from different universities around the globe. He served as a TPC member and invited reviewer for international journals and conferences.

...