SMOTE-ENN Based Deep Stacking Network Model for Heart Disease Prediction

Sonam Palden Barfungpa[1*], Leena Samantaray[2], Hiren Kumar Deva Sarma[3]
[1]Advanced Technical Training Centre, Bardang, East Sikkim, India
Research scholar, Biju Patnaik University of Technology, Odisha, India
[2]Ajay Binay Institute of Technology (ABIT), Markat Nagar, Cuttack Development Authority, Sector – 1, Odisha, India
[3]Department of Information Technology, Gauhati University, Jalukbari, Guwahati, Assam, India
Email: [1] palden73@gmail.com

Abstract
In healthcare field, cardiovascular disease prediction is the challenging task. It is a challenging task to predict cardiac disease with high accuracy. Recently, a number of ML-based systems for the forecast and analysis of heart illness have been introduced. However, systems can't manage large amount of datasets. To overcome above issues, a stacked ensemble net is proposed to provide highly accurate outcomes and enhances the efficiency of the performance without any high-dimensional dataset. Using a deep learning model, heart disease can be easily predicted, and doctors can diagnose heart disease and provide better treatment to the patients. The synthetic minority over-sampling technique and Edited nearest neighbor (SMOTE-ENN) accomplish oversampling and downsampling at the same time, and it balances the imbalanced dataset. The features are extracted from stacked DBN and stacked ANN, which is given to ensemble techniques such as decision tree (DT), bagging and K-Nearest Neighbors (K-NN). The recommended model is called a stacked ensemble net, which combines the stacked deep belief network (DBN) and stacked artificial neural network (ANN). It eliminates the possibility of data leaking and over fitting. The present work is estimated with cardio illness data and related to traditional classifiers on the basis of feature extraction and ensemble methods. Predictive methods such as DT, Bagging, K-NN, Adaboost, Gaussian naïve Bayes, Random forest, SVM and Logistic regression are estimated according to accuracy, recall, precision, and AUC. The proposed system acquires an accuracy of 95.8%, which is greater than state-of-art techniques. This study displays that, in contrast to other state-of-art techniques, the proposed work is more effective at predicting heart disease.
Keywords: Heart disease, stacked deep belief network (DBN), stacked artificial neural network (ANN), stacked ensemble net, decision tree (DT), bagging, k-nearest neighbour (K-NN).

1. Introduction

Heart disease is the primary reason of mortality and morbidity universally, contributing to above 70% of losses worldwide. As per 2017 Global Burden of Disease research, cardiovascular disease accounts for around 43% of all deaths. The World Health Organization (WHO) evaluates that over 17 million people pass away each year from cardiac illness, which is more frequent in the United States, Asia, and India. If the cardiac fails to activate effectively, it will affect other human body parts, like the kidneys and brain. Cardiac is just a pump that circulates blood throughout the human parts. Many body parts, including the cerebral cortex, suffer when there is inadequate oxygen in the body. Death occurs within minutes when the circulatory system stops beating. Everything about one’s life is determined by the heart’s capacity to work efficiently [1].

According to a medical study, lifestyle variables such as obesity, dietary habits, and lack of exercise are the main reasons of cardiac illness. Hypertension, Electrocardiogram rate, height, nicotine consumption, overweight, undue alcohol intake and diabetes can all raise the likelihood of developing heart disease [2, 3]. Preventing cardiovascular disease is vital, and an electrocardiogram (ECG) to predict cardiac infarction will allow to know in what way to identify constant and preventive process, ensuring that more people may live a healthy life. The ECG is a popular and simple method for diagnosing cardiovascular diseases. With the fast expansion of ECG testing and the scarcity of cardiac specialists, automated mining of features from ECG signals to aid in disease prediction has emerged as a popular research area. Computer-assisted approaches using ECG data are commonly employed for diagnosing or short-term prognosis of cardiovascular disorders and have shown promising results. However, an ECG cannot identify myocardial ischemia since it is frequently induced by stress. Forecasting the results of cardiovascular illnesses using activity ECG information is still a topic that requires more investigation [4].

Using deep learning and feature fusion techniques, it can predict heart disease. Deep learning approach to identify useful features and classify them correctly. Deep learning is used to extract and merge features from structured and unstructured patient data. In order to provide crucial health records, the feature merging approach first combines the obtained attributes from various sensors with Last but not least, the knowledge improvement approach eliminates pointless and terminated features while choosing the essential ones, reducing computational load and increasing system efficiency [5]. An ECG grayscale image and a scalogram image were created from the ECG signal from each lead, and these images were used to adjust the previously trained model. The stacking ensemble method used the model as a base learner. The single-modal stacked ensemble outperformed the 12 separate base learner ensembles and the single-modal simple averaging ensemble for both scalogram images and ECG grayscale images.

This type of decision-making system can help physicians detect diseases at their earliest stages. This study uses a stacked deep learning model to test the hopes of improving performance in heart disease diagnosis. The main objective patients with cardiac this study is to give physicians a tool to help in the premature problems. As an outcome, successfully treating patients and averting impacts will be simpler. To evaluate the performance of the proposed technique, 4 different parameters are applied: recall, F1-measure, accuracy, and precision.

The contributions of the presented study are: The outcomes of this study show the optimum modality for predicting or detecting adaptive metaheuristic-based deep learning techniques. It also helps academics and doctors develop the accuracy of cardiovascular disease. This comparative analysis of this study helps to identify the strengths and disadvantages of previously presented deep learning approaches for cardiovascular disease detection.

The organization of this paper is arranged as follows: Section 2 offers a brief related works of prediction of cardiovascular disease. Subsequently, Section 3 presents the proposed methodology and deliberates data pre-processing, SMOTE-ENN, extracting the features from the stacked DBN and ANN and ensemble techniques. Then, Section 4 provides the results and discussion. Finally, Section 5 concludes the proposed work and provides recommendations for further research.

2. Literature Survey

Liang et al. [6] recommended tBNA-PR, a patient illustration model based on a temporal Bidirectional neural network and an Attention mechanism deep learning model. tBNA-PR efficiently models diverse and dynamic temporal Electronic Health Records (tEHRs) data from the past and the future to produce useful patient depictions for precise heart failure diagnosis and fair patient stratification. In the context of complicated clinical settings, this research extracts common diagnoses and prescriptions for illness pattern investigation and discovers key elements of sub-phenotypes for subgroup description in order to give higher high-quality medical care and medical decision support. This study makes use of a practical dataset, the MIMIC-III database. The heart failure prediction studies examine tBNA-PR, which achieves an accuracy rate of 0.78, an F1 score of 0.7671, and an AUC of 0.7198.

Mohan et al. [7] developed a hybrid random forest with a linear model (HRFLM). The suggested HRFLM method is used to merge the characteristics of the Linear Method (LM) and (RF). The heart disease was taken from the Cleveland UCI repository. It gives a simple visual demonstration of the dataset, workplace conditions, and analytics for prediction. The ML method begins with data pre-processing and then proceeds to select features according to DT entropy, modelling performance categorization, and enhanced accuracy outcomes. The feature selection and modelling process is repeated for different attribute combinations. HRFLM demonstrated to be accurate in the forecast of heart illness.

Zhang et al. [8] suggested XGBoost and Random forest were applied as classification models. This is used to progress the performance and precision. Adaptive synthetic (ADASYN) and (SMOTE) were utilized to steadiness the dataset that has included 28.71% normal samples and 71.29% heart disease samples. The four feature extraction approaches increased classification model performance in categorization accuracy, precision, recall, F1 score, and specificity. On the set of data analyzed using feature creation and the SMOTE technique, the XGBoost algorithm produced the finest prediction outcomes for performance.

Krishnani et al. [9] progressed an intensive pre-processing strategy for predicting Heart Disease. Changing null values, standardization, resampling, normalization, prediction, and classification are part of an intensive pre-processing technique. This research is to forecast the risk of heart disease using machine learning methods like DT, RF and K-NN. Contrast analysis of these methods based on accuracy in predicting is also carried out. Furthermore, K-fold Cross Validation is employed to produce data randomization. These methods have been tested on the "Framingham Heart Study" dataset, which has 4240 records. K-NN, RF and DT attained an accuracy of 92.89%, 96.8%, and 92.7%, respectively, in the experimentation. As a result of pre-processing methods, Random Forest classification produces outcomes that are more precise than other kinds of machine learning algorithms.

Shorewala et al. [10] explored six approaches: Binary Logistic Classification, Naive Bayes and K-Nearest Neighbors. K-Fold validation will also generate unpredictability in the data and assess the reliability of the model's output. Additionally, hybrid models are investigated utilizing ensemble approaches such as bagging, boosting, and stacking. The outcomes of these ensemble approaches are contrasted with the outcomes of the initial basis classifiers. These algorithms are evaluated using the 'Heart Illness Dataset,' which has 70,000 medical examination data for heart illness. Compared to their conventional counterparts, bagged algorithms have an average enhanced accuracy of 1.96%. Boosted models obtained the greatest AUC score of 0.73 and overall accuracy of 73.4%. With an accuracy of 75.1%, the stacked model incorporating K-NN, RF, and SVM proved the most efficiency.

3. Proposed Methodology

This section briefly discussed the infrastructure of the proposed work in detail. Firstly, the pre-processing, SMOTE-ENN, feature extraction using stacked DBN and ANN is described. As a final point, the organization of the ensemble approaches are presented, which are engaged to forecast heart failure in patients and suggest nutritional plans and activities.

A. Pre-processing

Pre-processing is a technique for obtaining complete, consistent, and readable data. The quality of the data influences the mining obtained results by machine learning techniques. Quality data leads to quality decisions. Typically, data pre-processing entails a number of tasks related to text cleaning, formatting, and formation of its illustration for model training. Selection of clean data to allow the ML model to absorb significant data from a data set while estimating its results with fewer errors [12]. The pre-processing processes used with feature selection improve the accuracy of the classification techniques shown in Figure 1.

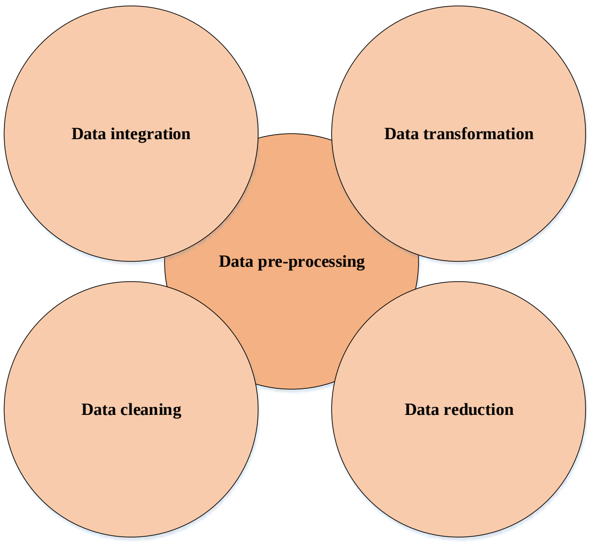

Figure 1: Pre-processing

Datasets can contain errors, missing data, redundant information, noise, and a host of other issues that render the data unsuitable for direct use by a machine learning algorithm. Data cleaning, data conversion, lack of imputation, data normalization, feature selection and other procedures depend on the type of data set [13].

Processing data is a crucial step during training. Subsequently, the model's ability to predict the future is greatly affected by the accuracy of the data. It can greatly improve the precision of a model. In addition, ensemble techniques achieve good results in the distribution of cardiovascular diseases [11].

B. SMOTE-ENN

SMOTE is an oversampling strategy that has been often used in treatment to cope with class unbalanced data. SMOTE enhances the amount of data instances through the generation of random, artificial data from its closest neighbors employing Euclidean distance. Since new instances are designed on the basis of original characteristics, they become same to the unique data. SMOTE may not be the greatest choice for managing high-dimensional data since it introduces additional noise. The SMOTE approach is used in this work to build a fresh training dataset [14]. The SMOTE-ENN equation (1) is given below,

$$X_{new} = x_i + (\hat{x}_i - x_i) \times \partial \tag{1}$$

The technique for SMOTE-ENN was employed for balancing the trained data distribution. The model was built using publicly available datasets (Kaggle), and its results were compared with that of different models GNB, DT, Bagging, K-NN, LR, SVM, RF, and ada boost [15].

C. Feature extraction using stacked DBN and ANN

In this proposed work, the balancing of the unbalanced data is performed using SMOTE-ENN. SMOTE rises the amount of data instances by producing random, artificial records from its nearest neighbors using Euclidean distance. It may not be the best choice for managing large datasets as it introduces additional noise. The technique for SMOTE-ENN was used to balance the trained data distribution. The model was built using publicly available datasets (Kaggle), and its results were compared to those of different models GNB, DT, Bagging, K-NN, LR, SVM, RF and Ada Boost.

K-NN offers high accuracy compared to ensemble techniques such as DT and bagging. K-NN offers high accuracy in predicting heart disease. Extracting the features from the collective using deep learning techniques such as Stacked DBN and Stacked ANN for high-precision cardiovascular disease prediction compared to other stacking ensemble methods such as DT, bagging and K-NN. Stacked DBN is used to mitigate underfitting and overfitting and increase the effectiveness of the method. DBN offers the expertise and accuracy of cardiovascular heart disease. Stacked ANN increases accuracy. The hybrid stacked DBN and ANN system is a stacked ensemble mesh offering high accuracy, precision, recall, etc.

1) Stacked DBN

DBN is a stochastic machine learning model comprising stacked Restricted Boltzmann Machine (RBM) modules. It is a system with numerous RBM intermediate layers and a classifier as the last layer of the network [16]. Figure 2 shows the architecture of DBN.

RBM are stacked to form the DBN distribution. To train multi-layer RBM without oversight, the architecture of the network first uses the Contrastive Divergence (CD) approach. Then the node variable is changed throughout the DBN structure using the BP method. DBN training mainly involves of pre-training and fine-tuning. In the pre-training stage, the CD approach is used to tune all different layers of RBM parameters while implementing unsupervised training with unlabelled sample data. Then, to achieve a global cluster weight improvement in the whole DBN network, the initial training DBN uses the BP technique to estimate the total network uncertainty of each individual layer and change the parameter values of all layers [17].

$$E_\theta(v,h) = -\sum_{i=0}^{n_v} a_i v_i - \sum_{j=0}^{n_h} b_j h_j - \sum_{i=0}^{n_v}\sum_{j=0}^{n_h} v_i w_{ij} h_j \tag{2}$$

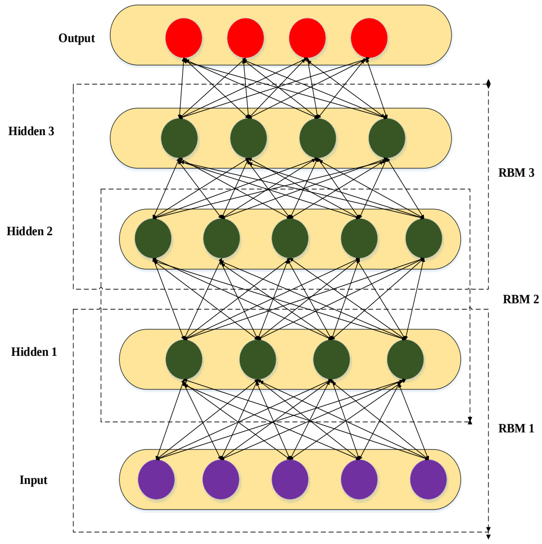

Figure 2: Architecture of DBN

2) Stacked ANN

A stacked ANN model was developed to improve the equity and security of infectious illness treatment. Using a method known as stacking, many ANNs may be combined to form an ML model. Stacked ANN is made up of 5 separate models of ANN that are employed to train a bigger ANN. Each of the models was trained using training with cross-validation, with every model trained on a different fold. Because ANNs are stochastic, every trained model has distinct weights, allowing it to learn its fold effectively. The last ANN develops from these several models, surpassing any particular model throughout the entire training set. ANN shown in Figure 3 compares the results from the relatively small ANNs using the test split and forecasts based on the best result.

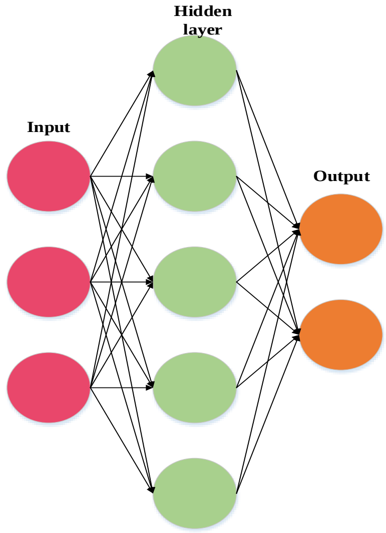

Figure 3: Architecture of ANN

The stacked ANN model is trained on a smaller sample than the ANN model, which is trained on the entire train split. The HD predicted by the Stacked-ANN model is acquired by means of the following equation (3):

$$\overline{HD}_{u,S-ANN} = \hat{g}(\overline{HD}_{u,RF}, \overline{HD}_{u,GB}, \overline{HD}_{u,SVM}, U_{u-1}, \ldots\ldots, U_{u-29}) =$$

$$= a^{(3)}\left(\sum_{m=1}^{10} x_{1,m}^{(3)} a^{(2)} \left(\sum_{n=1}^{20} x_{a,n}^{(2)} a^{(1)}\left(\sum_{k=1}^{33} x_{n,k}^{(1)} y_k + x_{n,0}^{(1)}\right) + x_{m,0}^{(2)}\right) + x_{1,0}^{(3)}\right) \tag{3}$$

The SANN model instead trains on the results of the individual smaller ANNs rather than directly learning from the entire train split. The altered train set is employed to train the ANN classifier, which subsequently predicts the test set. It eliminates the possibility of data leaking and over fitting [18].

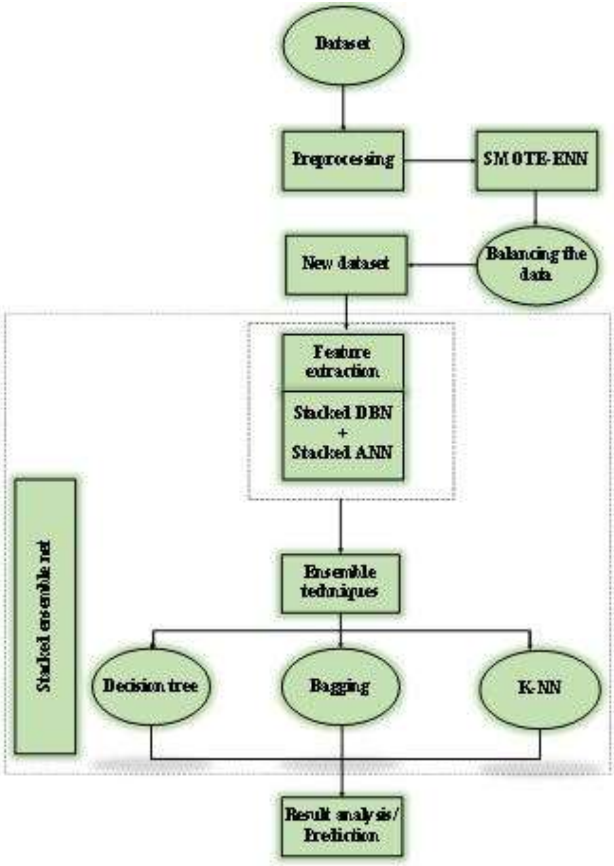

Figure 4: Proposed system flow

The proposed workflow is presented in Figure 4, where the dataset is pre-processed, and the unbalanced data is balanced by SMOTE-ENN. The extraction of features through a mixture of stacked DBN and stacked ANN. The proposed stacked ensemble techniques obtained better efficiency when associated to other existing techniques like DT, Bagging, K-NN, Adaboost, Gaussian naïve Bayes, Random forest, SVM, and Logistic regression.

D. ENSEMBLE TECHNIQUES

Stacking is achieved by combining three ensemble techniques, namely the DT, bagging, and the K-NN. An ensemble could perform better and make better predictions.

A model's performance and predictions' spread or dispersion are lessened by an ensemble [19].

1) Decision tree

DT are a popular classification technique that can be broken down into three versions based on the different ways trees are constructed. DT method also uses pre pruning and post pruning techniques to avoid overfitting and improve the system's ability to generalize [11]. The ML is a technique on the basis of supervised learning that constructs a DT from a series of training examples labelled with classes [20]. The following equation (4) is given below,

$$Entropy(E) = \sum_{k=1}^{1} -_{qk} \log_{2qk} \tag{4}$$

2) bagging

Bagging is the process of creating many copies of a forecaster and merging them to generate a consolidated forecaster. The aggregation aggregates several versions of the indicator through plurality voting. This improves the efficiency of a poor classifier by running parallel homogeneous classifiers and then combining the results with a function. The cross was contrasted with its original base model after each model was bagged [10].

3) K-NN

KNN is also a supervised classification method. It guesses the target class based on the comparability of the specific information with the other given information of the training tag algorithm. This can be viewed as a comparison of the properties of the information that is expected to be targeted with the properties of existing data [12]. The equation (5) of K-NN is given below,

$$P(x, x^i) = | x - x^i | = \sqrt{\sum_{i=1}^{d} (x_i - x_i^{'})^2} \tag{5}$$

KNN is a method of automated knowledge in which KNN trains after delivering test samples, and training samples incur no additional time cost [11].

4. Results & Discussion

The results of the performance for the ML classifiers such as DT, Bagging, K-NN, Adaboost, Gaussian naïve Bayes, Random forest, SVM, and Logistic regression are computed in Table 2 and examined the method's evaluations in Figure 5-9. Table 2 stated the outcomes of Accuracy, Precision, Recall, F1Score, and AUC for analyzing the methods evaluations. The dataset can be classified into two types: Training sets and test sets. After data pre-processing, SMOTE-ENN can balance the imbalanced data, stacked DBN and stacked ANN extracted the features, and it will be given to stacking ensemble techniques such as DT, bagging and K-NN. The proposed stacked ensemble net provides high accuracy and performs efficiently compared to other ensemble techniques.

A. Dataset

The data was retrieved from Kaggle and left unedited. The data were organized into categories and a file with comma-separated values. It contained no null values, and all variables were either continuous or categorical.

The data set showed extreme values, i.e. worldwide anomalies. To successfully address this issue and ensure data integrity, the upper and lower 2% percentiles were truncated for all continuous variables with a significant standard deviation. In addition, there were anomalies, such as cases where the systolic heart rate was lower compared to the distal heart rate. Finally, to ensure consistency across the data set, the unclassified numeric variables were normalized within a range of 0 and 1 [10]. Table 1 represented the Kaggle Datasets.

Table 1: Kaggle Datasets

Feature	Representation	Description
Age	Age	Patients age
Gender	Sex	Male=0;Female=1
Cardiac ache	Cp	4 types of chest pain (0-Typical angina,1-Atypical

		angina,2-nontypical angina,3-Asymtomatic)
Blood pressure	Bp	Min=120;Max=200
Blood sugar	Bs	Min=100;Max=200
Cholesterol	Chol	Min=100;Max=200
Heart rate	Thalach	Min=72;Max=100
Glucose	Gluc	Min=1;Max=3
Smoking	Smoke	Yes=1;No=0
Physical activity	Activity	Yes=1;No=0
Alcohol intake	Alco	Yes=1;No=0
Presence of heart disease	Cardio	Yes=1;No=0
Absence of heart disease	Cardio	Yes=1;No=0
Thalassemia	Thal	Normal=0;Abnormal=1

B. Performance analysis

The performance evaluation of the proposed stacked ensemble Net associated to other ensemble techniques such as DT, Bagging, K-NN, Gaussian Naive Bayes, SVM, Random Forest, Logistic Regression, Adaboost in terms of accuracy, precision, recall, F1 score and AUC.

Accuracy

Accuracy is determined as the proportion of precise forecasts through the algorithm to the overall amount of occurrences that is shown in equation (6).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}$$

Precision

Precision is determined as the amount of those who had been anticipated to be at risk of having cardiac illness.

$$precision = \frac{TP}{TP + FP} \tag{7}$$

Recall

The recall is the proportion of people at risk of getting heart disease and were projected to be at hazard by the algorithm. The formula for the recall is given below as equation (8)

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{8}$$

F1 Score

The harmonic average of recall and accurateness is the F1 Score [10] which is given below as equation (9)

$$F1Score = \frac{2(\mathrm{Pr}ecision \times \mathrm{Re}call)}{\mathrm{Pr}ecision + \mathrm{Re}call} \tag{9}$$

AUC

AUC is commonly employed to quantify the ROC curve. In other words, the nearer the AUC is to 1, the classifier's predictive accuracy is greater [11].

Table 2: Comparisons of ensemble techniques

Models	Accuracy (%)	Precision (%)	Recall (%)	F1Score (%)	AUC (%)
K-NN [10]	89.6	85.6	86.3	82.8	0.84
DT [10]	73.0	63.3	77.8	69.8	0.72
Bagging [21]	82.00	88.36	75.36	70.10	0.89

From the data available in Table 2, the ensemble techniques are compared to K-NN, which offers high accuracy, precision, recall, F1 score and AUC compared to other ensemble techniques such as DT and bagging.

Table 3: The outcomes of the ensemble techniques

Models	Accuracy (%)	Precision (%)	Recall (%)	F1Score (%)	AUC (%)
Stacked ensemble net	95.8	94	96.23	95	0.94
DT [10]	73.0	63.3	77.8	69.8	0.72
Bagging [21]	82.00	88.36	75.36	70.10	0.89
K-NN [10]	71.6	66.6	73.3	69.8	0.71
Gaussian Naïve Bayes [10]	71.4	76.3	61.2	67.9	0.70
SVM [10]	72.2	76.7	62.9	69.1	0.72
Random Forest [20]	61.70	43.46	37.50	39.77	0.64
Logistic Regression [20]	63.33	64.29	24.32	35.29	0.57
Ada Boost [10]	73.5	76.92	73.26	70.7	0.73

As presented in Table 3, the proposed model accomplished better efficiency with regard to accuracy, precision, recall, f1 score and AUC. The stacked ensemble net's performance is 95.8% in accuracy, 96.23% in recall, 94% in precision, 0.94% in AUC, and 95% in f1 score. The proposed model exhibits better efficiency when compared to other classifiers.

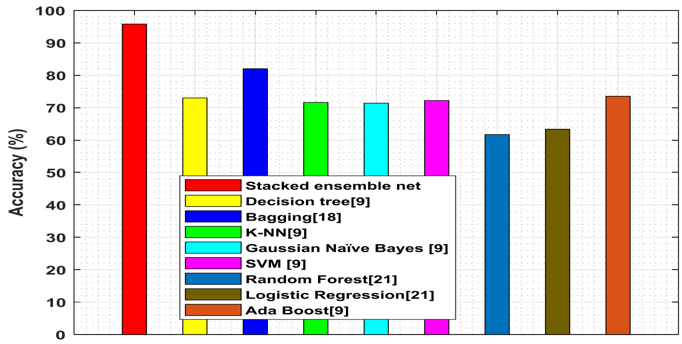

Figure 5: Comparisons of accuracy

Figure 5 shows a comparison of accuracy acquired from the proposed stacked ensemble net model with other individual classifiers like DT, Bagging, K-NN, Gaussian naïve Bayes, SVM, Random forest, Logistic regression, Ada boost. The proposed model's accuracy is 95.8% which is higher than other ensemble techniques. It provides a high accurate prediction of cardiovascular disease.

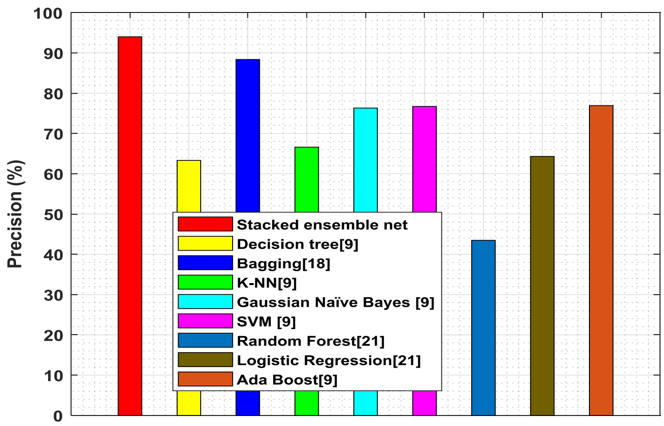

Figure 6: Comparisons of precision

Figure 6 illustrates a distinguish of precision acquired on proposed technique with other classifiers such as DT, Bagging, K-NN, Gaussian naïve Bayes, SVM, Random forest, Logistic regression, Ada boost. The proposed model's accuracy is 94% which is higher than other ensemble techniques.

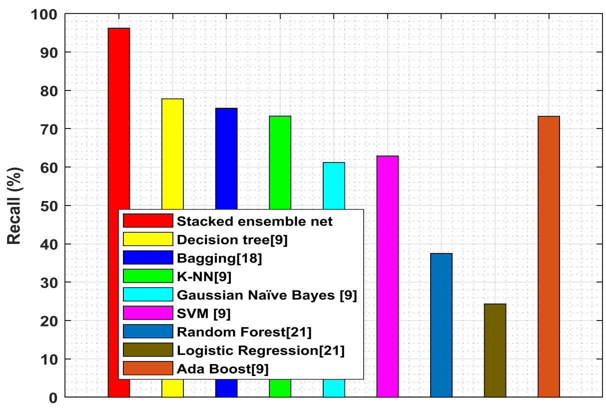

Figure 7: Comparisons of recall

Figure 7 demonstrates a differentiate of recall acquired on proposed system with other classifiers such as DT, Bagging, K-NN, Gaussian naïve Bayes, SVM, Random forest, Logistic regression, Ada boost. The proposed model's recall is 96.23% which is higher than other ensemble techniques.

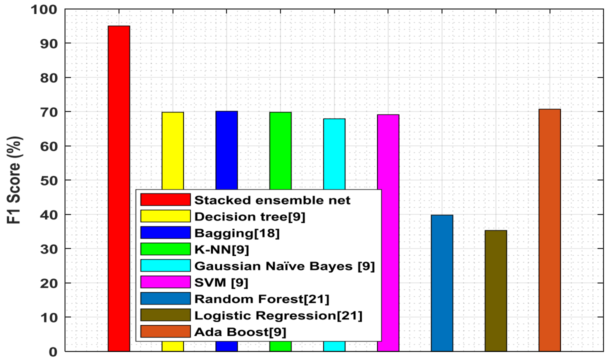

Figure 8: Comparisons of F1 Score

Figure 8 shows a comparison of the F1 Score acquired on recommended stacked ensemble net model with other distinct classifiers such as DT, Bagging, K-NN, Gaussian naïve Bayes, SVM, Random forest, Logistic regression, Ada boost. The proposed model's recall is 95%, which is higher than other ensemble techniques.

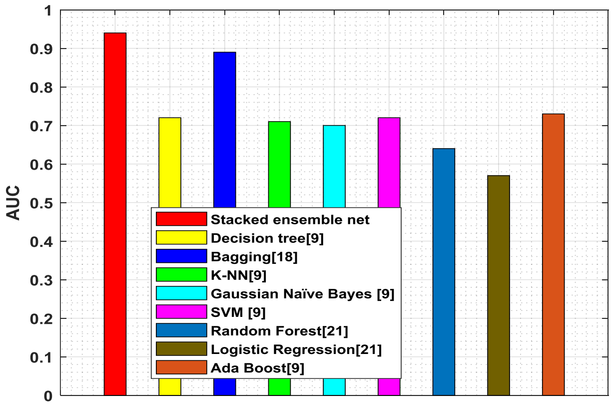

Figure 9: Comparisons of AUC

Figure 9 displays a contrast of recall acquired from the suggested stacked ensemble net technique with other distinct classifiers such as DT, Bagging, K-NN, Gaussian naïve Bayes, SVM, Random forest, Logistic regression, Ada boost. The proposed model's recall is 0.94% which is higher than other ensemble techniques.

5. Conclusion

In this paper, the investigation of imbalanced datasets indicates fewer exact forecasts unless the data set is properly balanced after pre-processing. Unless the data set is correctly balanced by pre-processing, analysis of unbalanced datasets results in less precise predictions. One of the most popular oversampling techniques for imbalanced datasets is SMOTE-ENN. A stacking strategy, including Stacked DBN and Stacked ANN, has been suggested to further boost accuracy. The stacking ensemble produced much greater accuracy as compared to individual classifiers, demonstrating a synergistic effect. Stacking ensemble net produced expressively higher accurateness. The individual classifiers of the DT, bagging, K-NN, Gaussian naïve Bayes, SVM, random forest, logistic regression, and ada boost showed accuracies of 73%, 82%, 71%, 71%, 72%, 61%, 63%, and 73%, although stacked DBN and ANN produced accurateness of 95% and 96%. The superior increase in accurateness is important meanwhile this offers an improved forecast of strength. The proposed model of stacked ensemble net of the accuracy, recall, precision, AUC and f1 score was 95.8%, 96.23%, 94%, 0.94%, and 95%, respectively. Future research will include ensemble techniques that can increase efficiency with many more system parameters for such techniques.

References

[1] C.M. Bhatt, P. Patel, T. Ghetia, and P.L. Mazzeo, “Effective Heart Disease Prediction Using Machine Learning Techniques”. Algorithms, vol. 16, no. 2, pp. 88, 2023.

[2] E.A. Ogundepo, and W.B. Yahya, “Performance analysis of supervised classification models on heart disease prediction”. Innovations in Systems and Software Engineering, pp.1-16, 2023.

[3] K.V.V. Reddy, I. Elamvazuthi, A.A. Aziz, S. Paramasivam, H.N. Chua and S. Pranavanand, “An Efficient Prediction System for Coronary Heart Disease Risk Using Selected Principal Components and Hyperparameter Optimization”. Applied Sciences, vol. 13, no. 1, pp.118, 2023.

[4] Y. You, W. Wang, D. Li, Y. Jia, D. Li, R. Zeng and L. Zhang, “Multi- modal machine learning based on electrocardiogram data for prediction of patients with ischemic heart disease”. Electronics Letters, vol. 59, no. 2, pp. e12708, 2023.

[5] F. Ali, S. El-Sappagh, S.R. Islam, D. Kwak, A. Ali, M. Imran and K.S. Kwak, “A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion”. Information Fusion, vol. 63, pp. 208-222, 2020.

[6] Y. Liang and C. Guo, “Heart failure disease prediction and stratification with temporal electronic health records data using patient representation”. Biocybernetics and Biomedical Engineering 2023.

[7] S. Mohan, C. Thirumalai and G. Srivastava, “Effective heart disease prediction using hybrid machine learning techniques”. IEEE access, vol. 7, pp.81542-81554, 2019.

[8] S. Zhang, Y. Yuan, Z. Yao, X. Wang and Z. Lei, “Improvement of the Performance of Models for Predicting Coronary Artery Disease Based on XGBoost Algorithm and Feature Processing Technology”. Electronics, vol. 11, no. 3, pp. 315, 2022.

[9] D. Krishnani, A. Kumari, A. Dewangan, A. Singh and N.S. Naik, “Prediction of coronary heart disease using supervised machine learning algorithms”. In TENCON 2019-2019 IEEE Region 10 Conference (TENCON) IEEE pp. 367-372, 2019, October.

[10] V. Shorewala, “Early detection of coronary heart disease using ensemble techniques”. Informatics in medicine Unlocked, vol. 26, pp.100655, 2021.

[11] J. Yang and J. Guan, “A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm”. Information, vol. 13, no. 10, pp. 475, 2022.

[12] S. Mondal, R. Maity, Y.R. Singh, S. Ghosh and A. Nag, “Early Prediction of Coronary Heart Disease using Boosting-based Voting Ensemble Learning”. In 2022 IEEE Bombay Section Signature Conference (IBSSC) pp. 1-5, 2022, December.

[13] M. Aljanabi, M.H. Qutqut and M. Hijjawi, “Machine learning classification techniques for heart disease prediction: a review”. International Journal of Engineering & Technology, vol. 7, no. 4, pp. 5373-5379, 2018.

[14] A. Ishaq, S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara and M. Nappi, “Improving the prediction of heart failure patients’ survival using SMOTE and effective data mining techniques”. IEEE access, vol. 9, pp. 39707-39716, 2021.

[15] N.L. Fitriyani, M. Syafrudin, G. Alfian and J. Rhee, “HDPM: an effective heart disease prediction model for a clinical decision support system”. IEEE Access, vol. 8, pp.133034-133050, 2020.

[16] P. Verma, V.K. Awasthi and S.K. Sahu, “Classification of Coronary Artery Disease Using Deep Neural Network with Dimension Reduction Technique”. In 2021 2nd International Conference for Emerging Technology (INCET) IEEE pp. 1-5, 2021, May.

[17] S.A. Alsuhibany, S. Abdel-Khalek, A. Algarni, A. Fayomi, D. Gupta, V. Kumar and R.F. Mansour, “Ensemble of deep learning based clinical decision support system for chronic kidney disease diagnosis in medical internet of things environment”. Computational Intelligence and Neuroscience, vol. 2021, 2021.

[18] B.U. Rindhe, N. Ahire, R. Patil, S. Gagare and M. Darade, “Heart Disease Prediction Using Machine Learning”. Heart Disease, vol. 5, no. 1, 2021.

[19] A.M. Sowjanya and O. Mrudula, “Effective treatment of imbalanced datasets in health care using modified SMOTE coupled with stacked deep learning algorithms”. Applied Nanoscience, vol. 13, no. 3, pp. 1829-1840, 2023.

[20] L.R. Guarneros-Nolasco, N.A. Cruz-Ramos, G. Alor-Hernández, L. Rodríguez-Mazahua and J.L. Sánchez-Cervantes, “Identifying the main risk factors for cardiovascular diseases prediction using machine learning algorithms”. Mathematics, vol. 9, no. 20, pp. 2537, 2021.

[21] M. Mamun, A. Farjana, M. Al Mamun, M.S. Ahammed and M.M. Rahman, “Heart failure survival prediction using machine learning algorithm: am I safe from heart failure?”. In 2022 IEEE World AI IoT Congress (AIIoT) pp. 194-200, 2022, June.