International Journal on Emerging Research Areas (IJERA) *ISSN:2230-9993*

DOI:

Prediction Model of Heart Disease for a Clinical Decision Support System

Athira A S
Computer Science and Engineering
(of KTU)
St.Joseph's College Of Engineering and Technology,SJCET
Kottayam, India
athirasudhakaran2423@gmail.com

Mr.Bino Thomas
Computer Science and Engineering
(of KTU)
St.Joseph's College Of Engineering and Technology,SJCET
Kottayam, India
bino.thomas@sjcetpalai.ac.in

Abstract—**Heart disease is one of the leading causes of death worldwide and can be mitigated by early diagnosis of heart disease. A clinical decision support system (CDSS) was used to diagnose heart problems earlier in the subjects. This paper proposes an heart disease prediction model (HDPM) for a CDSs consisting of Z score & Boxplot to detect and eliminate outliers, a hybrid synthetic oversampling technique edited by the nearest neighbors (SMOTE-ENN) to balance the distribution of training data and different machine learning algorithms predict and compare the model accuracy of heart disease. Build models using publicly available datasets (Statlog and Cleveland and hungary) and compare. The results show that the proposed model outperforms other models and previous research results . In addition, we designed and developed a prototype Cardiac CDSS (HDCDSS) to help physicians/clinicians diagnose heart problems in patients/subjects based on their current condition. Therefore, early treatment is possible to prevent death from too late diagnosis of heart disease.**

*Keywords—**Heart disease, Disease prediction model, Machine learning,Deep learning***

I. Introduction

Heart disease is one of the major diseases world wide.Many people are affected by heart disease every day. The causes of heart disease are high blood pressure, Heart disease is one of the major diseases world wide.Many people are affected by heart disease every day. The causes of heart disease are high blood pressure, high low-density lipoprotein (LDL) cholesterol, diabetes, smoking and secondhand smoke exposure, obesity, unhealthy diet, and physical inactivity. In modern days the Machine learning and deep learning Technologies are used to predict the heart disease status earlier. Machine learning (ML) proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry.Different types of machine learning and deep learning algorithms are used to diagnose heart disease.Heart disease is at the forefront of diseases in our society today. Heart disease is a cardiovascular disease (CVD) that remains the number one cause of death globally and contributes to approximately 30% of all global deaths[1].The total number of deaths globally is projected to increase to around 22 million in 2030.The main causes of heart disease are the over foods we eat, lack of exercise, smoking etc. It is better to treat the disease at the beginning than to treat it later. With today's modern facilities, heart disease can be diagnosed early. The technologies used for that are machine learning and deep learning. A variety of machine learning algorithms are used to predict heart

DOI:

disease.The all machine learning algorithms use the same process heart disease dataset collection, preprocessing, feature extraction, applied machine learning algorithms.The 14 parameters are used for the machine learning algorithm to predict the heart disease.The parameters are age,sex,chest-pain type,resting blood pressure,serum cholesterol,fasting blood sugar,resting ECG,max heart rate achieved,exercise induced angina,St depression induced by exercise relative to rest,peak exercise ST segment,number of major vessels colored by fluoroscopy,thal,diagnosis of heart disease.

The early heart disease identification of high-risk individuals and the improved diagnosis using a prediction model have generally been recommended to reduce the fatality rate and improve the decision-making for further prevention and treatment [5]–[7].A prediction model that is implemented in the clinical decision support system (CDSS) can be used to help clinicians assess the risk of heart disease and provide appropriate treatments to manage the risk further [8].In addition, numerous studies have also reported that the implementation of CDSS can improve preventive care, clinical decision making and decision quality [9]–[12].Machine learning-based clinical decision making have recently been applied in healthcare area. Previous studies have shown that machine learning algorithms (MLAs)such as chaos fly fireflyfly algorithm [13], backpropagation neural network (BPNN) [14], multilayer perceptron (MLP) [15],logistic regression (LR) [16], support vector machine (SVM) [17], and random forest (RF) [18] have been successfully used to help as decision making tools for heart disease prediction based on individual data.Several studies have also revealed the advantage of a hybrid model which achieved good performance in predicting heart disease such as majority voting of naïve bayes (NB), bayes net (BN), RF, and MLP [19], two stacked SVMs [20], and RF with a linear model [21].However, in the machine learning field, outlier and imbalance data may arise and impact on the performance of the prediction model.Previous studies have reported that by incorporating Density-Based Spatial Clustering of Applications with Noise (DBSCAN)-based to detect and eliminate the outlier data [22]–[24], and by balancing the distribution of data using a hybrid Synthetic Minority Over-sampling Technique-Edited Nearest Neighbor (SMOTE-ENN) [25]–[28], the prediction models' performances were significantly enhanced.To the best of our knowledge, no study has investigated a heart disease prediction model (HDPM) by utilizing, SMOTE-ENN and XGBoost machine learning.Therefore, we propose an effective HDPM for a CDSS which consists of Z Score & Boxplot-based to detect and eliminate the outliers, SMOTE-ENN to balance the training data distribution and XGBoost to predict heart disease. Our challenge is to detect and remove the outlier data and to balance the distribution of the training dataset to improve the performance of the HDPM. Two publicly available datasets (Statlog [29] andCleveland [30])Contributions of our study can be summarized as follows.

- proposed HDPM by integrating Z Score and Boxplot outlier detection, SMOTE-ENN, and different machine learning algorithms(random forest,logistic regression,support vector machine,Gussian naive bayes,decision tree,gradient boosting,k-nearest neighbor) improve prediction accuracy. The HDPM discovered from two public datasets and the trained version became applied to predict the subjects' coronary heart ailment reputation based on their modern-day situation.

- Overall performance evaluation and comparison with state-of-the arts fashions. The proposed HDPM becomes evaluated with other category models and as compared with the effects from preceding studies. They provided the statistical assessment to verify the enormousness of our model as compared to different fashions.

- Real case system improvement. designed and developed the prototype of the system to reveal the feasibility and applicability of our proposed model for actual-world

case look at. it's far anticipated that the developed device can be used as a sensible guiding principle for the healthcare practitioners.

The remainder of this study is organized as follows.Section II summarized the literature review. Section III presents the proposed HDPM including datasets description,overall design, and modules of the proposed model as well as performance evaluation metrics. Section IV discusses the performance evaluation of the proposed model, including the statistical test and comparison with previous studies. Section V presents the practical applications of the proposed model in the real case scenario. Finally, the concluding remarks and are presented in Section VI.

II . LITERATURE REVIEW

A heart disease prediction system's aim is to determine if a patient should be diagnosed with heart disease or not.Machine Learning can play an essential role in predicting presence/absence of heart diseases and more. Such information, if predicted well in advance, can provide

DOI:

important insights to doctors who can then adapt their diagnosis and treatment per patient basis.Several studies have reported the development of machine learning models for the diagnosis of heart disease. This learning occurs with training, where the algorithm develops an exercise in matching inputs to relevant outputs. Classification problems are a common framework for supervised learning tasks. Provides the purpose of HDPM. Many supervised machine learning algorithms are used to predict heart disease.Supervised learning algorithms include Logistic Regression (LR), Support Vector Machine (SVM), Random Forest, Naive Bayesian, Bayesian Network, Decision Tree, XGBoost

Islam et al. [2] proposed some superior data analysis techniques such as logistic regression In this case, LR provided the highest accuracy with 86.25%. technique for classification problems in machine learning. Logistic Regression is a little bit similar because both have the goal of estimating the values for the parameters or coefficients. After training a model, find out the relation between training and testing data. got 86.25% accurate results for logistic regression.S. M. Saqlain et al.[3] this paper, presenting a clinical heart disease diagnostic system by proposing a feature subset selection methodology with an object of achieving improved performance. The proposed methodology presents three algorithms for selecting candidate feature subsets: (1) mean Fisher score-based feature selection algorithm, (2) forward feature selection algorithm and (3) reverse feature selection algorithm. Liaqat ali et al.[4] this paper, introduces an expert system that stacks two support vector machine (SVM) models for the effective prediction of HF. The first SVM model is linear and L_1 regularized. It has the capability to eliminate irrelevant features by shrinking their coefficients to zero.The random forest algorithm has proven to be the most efficient algorithm for classification of heart disease.S. Dhanka et al.[5] This paper, Machine Learning (ML) is playing a substantial role in HD detection. Cardiologists, Physicians and Biomedical Engineers have collectively designed different ML algorithms to detect HD in initial stages. Ashir javeed et al.[6] This paper, learning system hybridizes two algorithms.The first algorithm is a random search algorithm which is used to search out subsets of features having complementary information about heart failure. The second algorithm was random forest which is used to predict heart failure based on the selected subset of features. It was shown that the proposed RSA-RF learning system improves the performance of random forest models by 3.3%. Additionally, the proposed learning system shows better performance than eleven recently proposed methods for heart failure detection and other well known machine learning models. It was also observed that the proposed system reduces the time complexity of the machine learning models by reducing the number of features. From the experimental results, we can conclude that the proposed learning system can help the physicians to improve the quality of heart failure detection.C. B. C. Latha et al.[7]this paper applies one such machine learning technique called classification for predicting heart disease risk from the risk factors.It also tries to improve the accuracy of predicting heart disease risk using a strategy termed ensemble..Shailaja et al.[8] Decision Support System helps in disease investigation by arranging a perception of health issues to the doctors or else by exposing background knowledge about individual patients. It additionally gives help with identifying a patient's situation and a guide to suggest the patient when to use the proper drug at a specific time, a web-based structure that has been involved with computerized patient/therapeutic records.Norma latif triyani et al.[9] this paper deals with the detection of heart disease.A clinical decision support system (CDSS) can be used to diagnose the subjects heart disease status earlier. This study proposes an effective heart disease prediction model (HDPM) for a CDSS

which consists Z score & Boxplot to detect and eliminate the outliers, a hybrid Synthetic Minority Over-sampling Technique-Edited Nearest Neighbor (SMOTE-ENN) to balance the training data distribution and XGBoost to predict heart disease.

Xiaoming Yuan et al.[10] this paper deals with Artificial intelligence (AI) that could be used for early prediction, detection, and diagnosis of health problems . It help patients obtain effective health guidance,intervention, and
treatment from doctors and could alleviate the serious influence of heart disease. Bo jin et al.[11] this paper deals with the Electronic health records (EHRs) containing patient diagnostic records, physician records, and records of hospital departments .This paper proposes a novel predictive model framework for heart failure diagnosis using LSTM methods.Compared to popular methods such as LR, RF, and AdaBoost,the method exhibits superior performance in the prediction of heart failure diagnosis. In the experimental data analysis and preprocessing, they used one-hot encoding and word embedding vectors to represent the patient diagnostic events.By analyzing the results, we reveal the importance of respecting the sequential nature of clinical records.

Wenbin chang et.al.[12] This paper presents a new hybrid XGBSVM model. To validate performance, the model was used to predict whether hypertensive patients develop hypertensive heart disease within three years. There are two evaluation indicators; one indicator is the traditional AUC, and the other indicator is the one proposed in this paper. The XGBSVM is compared with five other models. The final experimental results prove that XGBSVM is highly feasible, stable and accurate, and it is a new model that can be applied practically. In the experiment, the final output of this model is the probability of patients being classified into positive categories.senthilkumar et al.[13] This paper deals with the hybrid HRFLM approach used combining the characteristics of Random Forest (RF) and Linear MethodLM). HRFLM

DOI:

proved to be quite accurate in the prediction of heart disease.accuracy level of 88.7% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM)Liaqat Ali et. al[14]this paper deals with the use of the χ2statistical model while the optimally configured deep neural network (DNN) is searched by using exhaustive search strategy. The strength of the hybrid model named χ2 -DNN is evaluated by comparing its performance with conventional ANN and DNN models,This model achieves a prediction accuracy of 93.33%.Binhu Wang et al.[15]this paper deals with the Renal dysfunction, which is associated with bad clinical outcomes, is one of the most common complications of heart failure (HF). Timely prediction of renal dysfunction can help medical staff intervene early to avoid catastrophic consequences. In this paper, multi-task deep and wide neural networks (MT-DWNN) for predicting fatal complications during hospitalization.The experimental results that the MT-DWNN model achieves better prediction performance on renal dysfunction in HF patients than conventional models.M. Raihan et al.[16] Ischemic Heart Disease (IHD) is a cardiovascular disease. About 6.96% people died from IHD . For some reason like smoking, drinking alcohol, stress and unhealthy food]. If proper precaution can be taken, the rate of IHD will decrease. To decrease the mortality rate and reduce the diagnosis cost by gaining maximum accuracy. ANN can be used to achieve efficient diagnosis results. In the recent era, due to some change in the environment and unhealthy food many people suffer from various diseases.This paper tried different methods for the risk prediction but got the best results using the ANN concept to predict IHD. In this method, the backpropagation algorithm with multilayered perceptrons has been implemented to predict the risk of IHD.

III. *MATERIALS AND METHODS*

The HDPM concept aims to provide predictive performance in the presence or absence of heart disease based on the patient's current condition. Flowchart in Figure 1 shows how HDPM is configured. First, gather heart disease information. Next, data preprocessing and feature selection of the data transfer are performed. Third, usingZscore,Boxplot-based outlier detection method to find outliers provides the best parameter .Fourth, the detected outliers are then subtracted from the training data. Fifth, evaluation data based on the SMOTE-ENN method is used to balance the training data. Random forest ,Logistic regression,Gaussian naive bayes,Decision Tree,Gradient Boosting,k-nearest Neighbour MLA is used to learn and generate HDPM from training data. Finally, a performance test is provided to evaluate the performance of the proposed model and the resulting HDPM is then used in CDSS. In our study, we used a 10-fold cross-validation method to avoid overfitting.Cross validation allows the model to learn from different training datasets by resampling; thus complementing the data used for validation and helps prevent overloading. Previous studies have shown that 10-fold cross-validation can be used to control for variance-variance trade-off, ultimately providing an overall model and preventing overfitting.

A. *.DATASETS*

To evaluate the performance of various heart disease prediction methods, different type data sets are used to help the heart disease prediction.

Statlog and Cleveland

It uses heart disease datasets (Statlog and Cleveland,hungary) to investigate how heart disease can be identified by applying the machine learning model. The proposed model is then applied to those two datasets and with the expectation of providing a general and robust HDPM.The University of California Irvine (UCI) Repository Statlog Heart Disease database website presents a dataset to investigate heart disease [17]. The original dataset consists of 270 subjects, 13 attributes and one output class (120 and150 subjects are labeled with the presence (positive class)and absence (negative class) of heart disease, respectively).There are no missing values in the dataset . A detailed attributes description (including data type and range) and distribution (mean and standard deviation (STD)) Dr. Robert Detrano, M.D., provided dataset II (Cleveland Heart Disease dataset) to investigate heart disease that was collected from the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation in California, United States [18].The original dataset comprises 303 subjects and used heart disease datasets (We used two heart disease datasets (Statlog and Cleveland) to investigate how heart disease can be identified by applying the machine learning model. The proposed model is then applied to those three datasets and with the expectation of providing a general and robust HDPM.The University of California Irvine (UCI) Repository Statlog Heart Disease database website presents a dataset to investigate heart disease . The original dataset consists of 270 subjects, 13 attributes and one output class (120 and150 subjects are labeled with the presence (positive class)and absence (negative class) of heart disease, respectively).There are no missing values in the dataset . A detailed attributes description (including data type and range) and distribution(Dr. Robert Detrano, M.D., provided a dataset (Cleveland Heart Disease dataset) to investigate heart disease that was collected from the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation in California, United States [18]. The original dataset comprises 303 subjects and ;termed datasets and to investigate how heart disease can be identified by applying the machine learning model. The proposed model is then applied to those two datasets and with the expectation of providing a general and robust HDPM.The University of

DOI:

California Irvine (UCI) Repository Statlog Heart Disease database website presents a dataset to investigate heart disease [17]]. The original dataset consists of 270 subjects, 13 attributes and one output class (120 and150 subjects are labeled with the presence (positive class)and absence (negative class) of heart disease).There are no missing values in the dataset . A detailed attributes description (including data type and range) and distribution .Dr. Robert Detrano, M.D., provided a dataset (Cleveland Heart Disease dataset) to investigate heart disease that was collected from the V.A. Medical Center, LongBeach and Cleveland Clinic Foundation in California, UnitedStates [18]. The original dataset comprises 303 subjects and 79 raw attributes, although only 13 attributes are used, and one attribute as an output class. We removed 6 subjects' data due to missing values and used the remaining 297 data in the pre-processing stage. The original class value is a multi-class variable with the value range from 0 to 4. The 0 value is used to represent the absence of heart disease while the values from 1 to 4 are used to represent the presence of heart disease with its stage condition.

Hungary, Switzerland, and Long Beach V

The Datasets for both Hungary, Switzerland, and Long Beach V and UCI Kaggle are used to analyze models.The dataset utilized from [19] which is a shortened version of [20]. The dataset contains 76 characteristics, including the class attribute, for 1025 patients from Cleveland, Hungary, Switzerland, and Long Beach V, however only a subset of 14 are used in this study. The heart disease prediction dataset has 14 columns, with 13 independent components and one dependent target variable. The target variable is separated into two groups: those who have heart diseases and those who do not. There are 1025 rows in all for 918 patients from the heart disease UCI Kaggle dataset, however only a subset of 12 are used in this study. The heart disease prediction dataset has 12 columns, with 11 independent components and one dependent target variable. The target variable is separated into two groups: those who have heart disease and those who do not. There are 918 rows in all. There are no missing values in the dataset.There are no missing values in the dataset.an in depth attributes description (inclusive of information kind and variety) and distribution(imply and preferred deviation (STD)) for dataset are given in Table 1.

B. Z SCORE & BOXPLOT*(outlier detection)*

Z-rating is a statistical measurement that describes a value's courting to the imply of a collection of values. Z-score is measured in terms of widespread deviations from the suggested. If a Z-rating is zero, it shows that the facts point's score is the same as the implied score. A Z-score of one.0 would suggest a price that is one trendy deviation from the suggested. Z-scores can be high-quality or bad, with a wonderful fee indicating the rating is above the suggest and a negative score indicating it's far underneath the imply.In making an investment and trading, Z-ratings are measures of an instrument's variability and can be used by traders to assist determine volatility. The Z-rating is from time to time stressed with the Altman Z-score, that's calculated using factors taken from an enterprise's monetary reviews. The Altman Z-score is used to calculate the probability that a commercial enterprise will move bankrupt inside the subsequent two years, while the Z-rating can be used from time to time stressed with the Altman Z-score, that's calculated using factors taken from an enterprise's monetary reviews. The Altman Z-score is used to calculate the probability that a commercial enterprise will move bankrupt inside the subsequent two years, while the Z-rating can be used to determine how far a stock's return differs from its average return—and lots extra.Z-rating is simply every other shape of preferred deviation process. Z-rating is used to convert the records into another dataset with imply = 0.

$$z = \frac{x - \mu}{\sigma}$$

$$\mu = \text{Mean}$$
$$\sigma = \text{Standard Deviation}$$

Z-score, that's calculated using factors taken from an enterprise's monetary reviews. The Altman Z-score is used to calculate the probability that a commercial enterprise will move bankrupt inside the subsequent two years, while the Z-rating can be used to determine how far a stock's return differs from its average return—and lots extra.Z-rating is simply every other shape of preferred deviation process. Z-rating is used to convert the records into another dataset with imply = 0.

right here, X-bar is the implied cost and s is trendy deviation. As soon as the facts are converted, the center becomes zero and the z-score similar to each fact factor represents the space from the middle in terms of widespread deviation. For example, a z-score of 2.5 shows that the statistics point is 2.5 trendy deviation away from the imply. typically z-rating =3 is taken into consideration as a cut-off cost to set the restrict. Consequently, any z-rating extra than +3 or less than -three is taken into consideration as an outlier that is pretty much just like the preferred deviation technique.

DOI:

Attributes			
Symbol	**Description**	**Type**	**Data Range**
age	Subject age in years	Numeric	[29, 77]
sex	Subject gender	Binary	0 = female, 1 = male
cp	Chest pain type	Nominal	1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, 4 = asymptomatic
trestbps	Resting blood pressure in mmHg	Numeric	[94, 200]
chol	Serum cholesterol in mg/dl	Numeric	[126, 564]
fbs	Fasting blood sugar with value > 120 mg/dl	Binary	0 = false, 1 = true
restecg	Resting electrocardiographic result	Nominal	0 = normal, 1 = having ST-T wave abnormality, 2 = showing probable or definite left ventricular hypertrophy
thalach	Maximum heart rate	Numeric	[71, 202]
exang	Exercise induced angine	Binary	0 = no, 1 = yes
oldpeak	ST depression induced by exercise relative to rest	Numeric	[0, 6.2]
slope	Slope of the peak exercise ST segment	Nominal	1 = up-sloping, 2 = flat, 3 = down-sloping
ca	Number of major vessels (0-3) colored by flourosopy	Nominal	0 – 3
thal	Defect type	Nominal	3 = normal, 6 = fixed defect, 7 = reversable defect

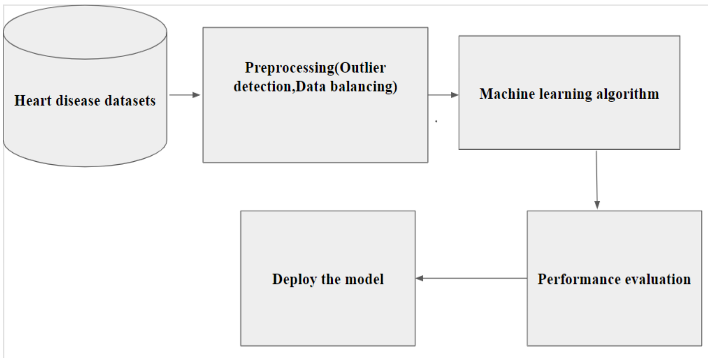

Fig. 1 :Heart disease prediction model of a heart disease clinical decision system

and the median of the dataset.1/3 quartile (Q3/seventy fifth Percentile): the middle price among the median the best value (no longer the "maximum") of the dataset.InterQuartile range (IQR): 25th to the 75th percentile. IQR tells how the middle values are."maximum": Q3 + 1.five*IQR"minimal": Q1 -1.5*IQROutliers: (shown as inexperienced circles) In facts, an outlier is a commentary point that is distant from other observations.

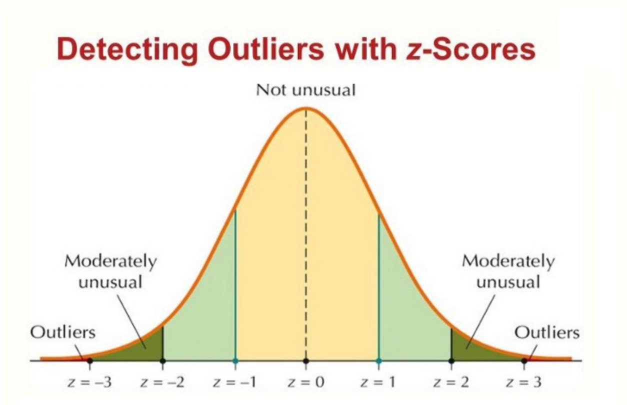

Figure 2: Detecting Outlier with z score

C . SMOTE-ENN-BASED DATA BALANCING

Information sampling or facts balancing is a not unusual method consisting of three subcategories, over-sampling, undersampling, and hybrid approach, and is utilized in device learning to address imbalanced data. determine 7 illustrates the 3 subcategories of statistics balancing strategies. The over-sampling technique balances the education data by means of generating facts samples beneath-sampling achieves that g oal with the aid of doing away with the information samples in the majority class.

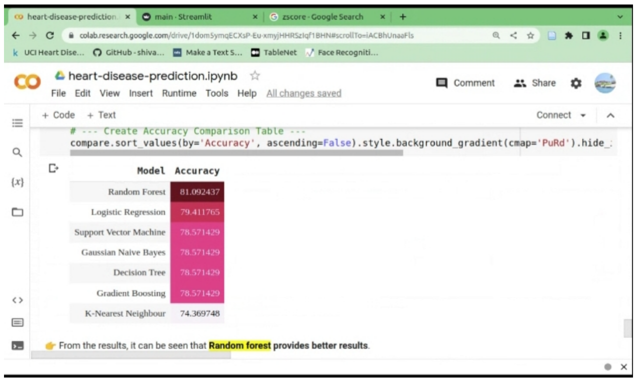

Figure 3:Model accuracy for different machine learning algorithms

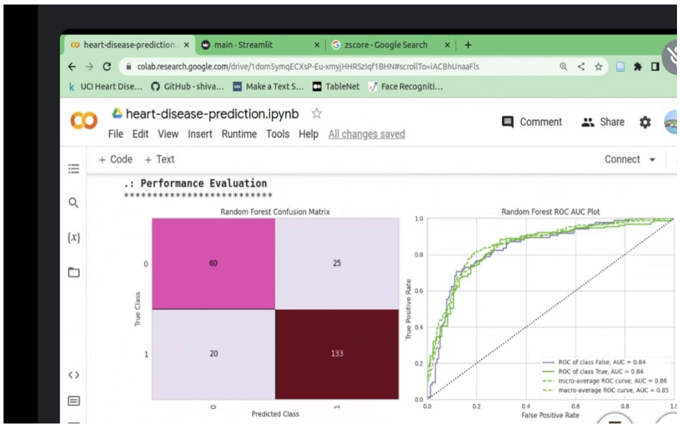

Figure 4:performance evaluation random forest confusion Matrix

DOI:

In the meantime, the hybrid method achieves the balanced facts by way of combining the over-sampling and beneath-sampling strategies.We used a hybrid SMOTE-ENN method to stabilize the imbalance heart sickness education datasets. In general,SMOTE is used to over-sample the minority magnificence until the education dataset is balanced, then the Edited Nearest Neighbor (ENN) is used to remove the unwanted overlapping samples between two classes while keeping the balanced distributions..

for the minority class at the same time as the beneath-sampling achieves that goal with the aid of doing away with the information samples in the majority class.

In the meantime, the hybrid method achieves the balanced facts by way of combining the over-sampling and beneath-sampling strategies.We used a hybrid SMOTE-ENN method to stabilize the imbalance heart sickness education datasets. In general,SMOTE is used to over-sample the minority magnificence until the education dataset is balanced, then the Edited Nearest Neighbor (ENN) is used to remove the unwanted overlapping samples between two classes while keeping the balanced distributions.

IV. RESULTS AND DISCUSSIONS

PERFORMANCE EVALUATION OF PROPOSED HDPM

According to the study, the proposed High-Dimensional Predictive Model (HDPM) performed better in terms of prediction accuracy when compared to other commonly used machine learning algorithms such as Naive Bayes (NB), Logistic Regression (LR), XGBoost, Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF).

The research team evaluated the performance of these algorithms based on eight different metrics, including accuracy, precision, recall/sensitivity/true positive rate, f-measure, Matthews correlation coefficient (MCC), false positive rate, false negative rate, and true negative rate. The results showed that the HDPM model outperformed the other algorithms in all of these metrics.

It's worth noting that the algorithms used in this study have been extensively studied and validated by the research community. The superior performance of HDPM suggests that it could be a promising approach for predictive modeling in high-dimensional datasets.

In conclusion, the HDPM model demonstrated superior performance compared to other commonly used machine learning algorithms, which could lead to more accurate predictions in high-dimensional datasets.The results of the comparative study indicate that the Random Forest algorithm performs the best among the machine learning algorithms tested. The accuracy achieved by the Random Forest algorithm is notably high.

V. CONCLUSION

This paper presents a comprehensive survey of the literature on Heart Disease Prediction. The survey covers the challenges and potential applications in the field, and classifies and analyzes recent methods while discussing various algorithms.

To structure the survey, the authors have organized the application scenarios into different categories in the Heart Disease Prediction Model (HDPM) literature. The survey provides valuable insights into the use of HDPM in different contexts, such as in hospitals and for the benefit of doctors and clinicians.

Overall, the Heart Disease Prediction system has the potential to be a very useful tool in the healthcare sector. It can help doctors and clinicians make informed decisions by providing accurate predictions about heart disease. As the number of heart disease cases continues to increase, the use of HDPM can play a critical role in improving patient outcomes and reducing healthcare costs.

References

1. World Health Organization. (2017). *Cardiovascular Diseases (CVDs)*.[Online]. Available:https://www.who.int/health-topics/cardiovasculardisease

2. Islam, S., Jahan, N., & Khatun, M. E. 2020. CardiovascularDiseaseForecast using Machine Learning Paradigms.2020 Fourth International Conference on Computing Methodologies and Communication(ICCMC). doi:10.1109/iccmc48092.2020.iccmc-00091.

3. .S. M. Saqlain, M. Sher, F. A. Shah, I. Khan, M. U. Ashraf, M. Awais,and A. Ghani,2019 ''Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines,'' Knowl. Inf. Syst., vol. 58, no. 1, pp. 139–167.

4. Liaqat ali ,awais niamat,javed ali khan,, noorbakhsh amiri golilarz , xiong xingzhong , adeeb noor , redhwan nour , and syed ahmad khan bukhari,2019,"An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure" in IEEE Access, vol. 7, pp. 54007-54014, 2019, doi: 10.1109/ACCESS.2019.2909969.

5. S. Dhanka and S. Maini,"Random Forest for Heart Disease Detection: A Classification Approach, 2021,"IEEE 2nd International Conference On Electrical Power and Energy Systems (ICEPES), Bhopal, India, pp. 1-3, do 10.1109/ICEPES52894.2021.9699506.

6. .A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor and R. Nour,2019 "An Intelligent Learning System Based on Random Search Algorithm and

DOI:

Optimized Random Forest Model for Improved Heart Disease Detection," in *IEEE Access*, vol. 7, pp. 180235-180243, 2019, doi: 10.1109/ACCESS.2019.2952107.

7. C. B. C. Latha and S. C. Jeeva, 2019,‘‘Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques,’’ *Inform. Med. Unlocked*, vol. 16, Jan. 2019, Art. no. 100203, doi:10.1016/j.imu.2019.100203.

8. Shailaja, K., Seetharamulu, B., & Jabbar, M. A., 2018. “Machine Learning in Healthcare”: A Review. Second International Conference on Electronics,Communication and Aerospace Technology (ICECA). doi:10.1109/iceca.2018.8474918.

9. .N. L. Fitriyani, M. Syafrudin, G. Alfian and J. Rhee,2020, "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System," in IEEE Access, vol. 8, pp. 133034-133050, doi: 10.1109/ACCESS.2020.3010511.

10. X. Yuan, J. Chen, K. Zhang, Y. Wu and T. Yang, "A Stable AI-Based Binary and Multiple Class Heart Disease Prediction Model for IoMT," in *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 2032-2040, March 2022, doi: 10.1109/TII.2021.3098306.

11. .B. Jin, C. Che, Z. Liu, S. Zhang, X. Yin and X. Wei, 2018,"Predicting the Risk of Heart Failure With EHR Sequential Data Modeling," in IEEE Access, vol. 6, pp. 9256-9261, doi: 10.1109/ACCESS.2017.2789324.

12. W. Chang, Y. Liu, X. Wu, Y. Xiao, S. Zhou and W. Cao,2019, "A New Hybrid XGBSVM Model: Application for Hypertensive Heart Disease," in IEEE Access, vol. 7, pp. 175248-175258, doi: 10.1109/ACCESS.2019.2957367.

13. S. Mohan, C. Thirumalai and G. Srivastava,2019, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in *IEEE Access*, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS..2923707.

14. L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed and J. A. Khan, 2019,"An Automated Diagnostic System for Heart Disease Prediction Based on ${\chi^{2}}$ Statistical Model and Optimally Configured Deep Neural Network," in IEEE Access, vol. 7, pp. 34938-34945, doi: 10.1109/ACCESS.2019.2904800.

15. B. Wang, Yongyi bai , Zhenjie yao, Jiangong li , wei dong yanhui tu, wanguo xue ,yaping tian ,yifei wang and kunlun HE ,2019, "A Multi-Task Neural Network Architecture for Renal Dysfunction Prediction in Heart Failure Patients With Electronic Health Records," in *IEEE Access*, vol. 7, pp. 178392-178400, doi: 10.1109/ACCESS.2019.2956859.

16. M. Raihan ,M. Raihan1 , Parichay Kumar Manda , Muhammad Muinul Islam , Tanvir Hossain, Promila Ghosh , Shakil Ahmed Shaj, Abdullah Anik , Mubtasim Rafid Chowdhury , Saikat Monda and Arun More,2019,"Risk Prediction of Ischemic Heart Disease Using Artificial Neural Network,"International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-5, doi: 10.1109/ECACE.2019.8679362.

17. *Statlog (Heart) Data Set*. Accessed: Oct. 2, 2019. [Online]. Available:http://archive.ics.uci.edu/ml/datasets/statlog+(heart)

18. *Heart Disease Data Set*. Accessed: Oct. 2, 2019. [Online]. Available:https://archive.ics.uci.edu/ml/datasets/Heart+Disease

19. [Online].Available: https://www.kaggle.com/johnssmith88/heart-diseasedataset

20. [Online]. Available: https://archive.ics.uci.edu/ml/dataset/heart+disease