Vol. 06, No. 4 (2024) 1871-1878, doi: 10.24874/PES.SI.25.03A.017

Proceedings on Engineering Sciences

www.pesjournal.net

HEART DISEASE PREDICTION USING MACHINE LEARNING CLASSIFIERS WITH VARIOUS BALANCING TECHNIQUES

Uzama Sadar
Parul Agarwal[1]
Suraiya Parveen
Geetika Dhand
Kavita Sheoran

Received 05.01.2024.
Received in revised form 08.02.2024.
Accepted 11.03.2024.
UDC – 004.85

Keywords:

Machine Learning, Heart Disease, Prediction Model, Balancing Techniques, Classification Algorithms

A B S T R A C T

Heart disease or Cardiovascular illness is the most prevalent cause of mortality globally. The challenge of predicting heart illness using clinical data analytics is considerable. Machine learning (ML) has been extensively used in the medical domain for disease prediction. This work performs a comparative analysis of various oversampling methods like the Synthetic Minority Oversampling technique (SMOTE), Synthetic Minority Oversampling Technique with Edited Nearest Neighbor (SMOTE-ENN), and Adaptive Synthetic Sampling Approach (ADASYN) Algorithm used with ML classifiers on imbalanced heart failure prediction dataset. Six ML classifiers are analyzed in the study Logistic Regression (LR), Support Vector Machine (SVM), K Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), and Gradient Boosting (GB). The accuracy metric is used to measure the model's performance The result depicts that the ADASYN technique performs better for the given dataset and increases the accuracy of the classifier in heart failure prediction.

1. INTRODUCTION

The heart is considered the central organ of a human being. Cardiovascular Disease (CVD) has been described as the most severe and fatal disease worldwide. A significant risk and burden are being placed on the world's healthcare system by the rise in cardiovascular illness with a high death rate.

A World Health Organization (WHO) research states that 17.9 million individuals worldwide die from CVD every year (World Health Organisation, 2020). If appropriate action is not taken, that number will rise to 22 million in 2030 (Nagavelli, Samanta & Chakraborty, 2022). CVD is a disease of the blood vessels and Heart. Blockage of arteries that supply oxygen and blood causes heart disease. According to the report NHIS (National Health Interview Survey), the five most prevailing symptoms of Heart attack are (Fang, Luncheon, Ayala, Odom, & Loustalot, 2019) Chest pain, Excessive Sweating, Dizziness, Fatigue, and Pain in the arm and jaw. Angiography is still the most reliable technique for identifying cardiac anomalies, but it

[1] Corresponding author: Parul Agarwal
Email: pagarwal@jamiahamdard.ac.in

is costly and requires a high degree of technical competence (Phegade et al., 2019) Therefore, clinicians are supported by a Machine learning-based predictive decision support system to reduce the death rate and improve the decision-making process (Rani, Kumar, Ahmed & Jain, 2021). Nowadays, Artificial Intelligence subparts Machine Learning and Deep Learning have been extensively used in clinical decision-making and assist in disease detection and prediction (Ramana et al., 2022). With the development of computers, the healthcare industry now has access to boundless amounts of clinical data in MRI, sensor, and electronic health record data (Dhand, Sheoran, Agarwal & Biswas, 2022) It is challenging for medical professionals to extract pertinent information from this data.

1.1 The paper's primary contribution is explained as follows

- This work performs a comparative analysis between oversampling methods.
- Three oversampling methods namely SMOTE, SMOTE-ENN, and ADASYN are integrated with 6 states of the art algorithm LR, SVM, KNN, DT, RF, and GB.
- The experimentation was done on the imbalanced Heart Failure Dataset.
- Accuracy metric was used for performance evaluation and results show that ADASYN performs better for the given dataset.

The remaining part of the paper is structured as follows: Section 2 describes related work with a brief description and the proposed methodology Section 3 discusses materials and methods including dataset description, classification techniques, and sampling techniques. Section 4 discusses the experimental results. Conclusion and future work are presented in Section 5.

2. LITERATURE SURVEY

Heart Disease commonly known as cardiovascular disease, is one of the worst diseases that contributes significantly to global mortality. Numerous researchers have put much effort into classifying and predicting CVD (Abdellatif et al., 2022). Several researchers have suggested Machine Learning (ML) algorithms to improve disease prediction accuracy. Many studies have thoroughly evaluated the presence of missing values in the dataset an important process, to increase the precision of results. In the paper authors (Gupta, Kumar, Arora & Raman, 2020) replaced missing values in the Cleveland dataset using the Pearson Correlation Coefficient along with the ML classifiers. Additionally, the Feature Selection approach is key to increasing the model's accuracy (Guyon & Elisseef, 2003). In the paper (Shah et al., 2017) authors used Principal Component analysis (PCA) to select features in heart disease prediction. Similarly in the paper (Khadir & Amanullah, 2018) to select features authors used Particle swarm Optimization (PSO). The proposed Support Vector Machine (SVM) with PSO is used to select significant features and achieves an accuracy of 99.7%. In the paper (Garate-Escamila, Hassani, & Andres, 2020), the authors enhanced the prediction accuracy of ML classifiers by combining CHI square with PCA. For Cleveland and Hungarian Datasets Random Forest (RF) classifier achieves the maximum accuracy of 98.7% and 99%. Hybrid models, which combine several ML models offer a potential method for disease prediction. The authors proposed A Heart Disease Prediction Framework using a Hybrid classifier and Genetic Algorithm in the paper (Ashri, EL-Gayar, & El-Daydamony, 2021). From the UCI repository cardiovascular dataset is used for experimentation and the proposed model attains an accuracy of 98.18%. Likewise in this paper (Mohan, Thirumalai, & Srivastava, 2019), the authors developed a novel prediction model using feature combinations and various classification techniques. The proposed hybrid Random Forest and Linear method achieves an accuracy of 88.7%. In the work (Tama, Im, & Lee, 2020) authors performed two-tier ensemble coronary disease detection using RF, Gradient boosting, and extreme gradient boosting. The model achieves an accuracy of 98.13%, and F1 of 96.6% respectively. However, the aforementioned works have certain limitations in detecting and predicting heart disease using the offered approaches because of the imbalanced clinical datasets.

Few works addressed imbalanced problems and developed a decision support system by utilizing balancing techniques. In the paper (Fitriyani, Syafrudin, Alfian, & Rhee, 2020), the authors proposed a model for heart disease prediction by combining Density-Based Spatial clustering (DBSCAN) of application with noise and Synthetic Minority Oversampling edited Nearest Neighbour along with extreme gradient Boosting (SMOTE-ENN) and achieving an accuracy of 98.4% for the Cleveland and 95.9% for Statlog dataset respectively. In the paper (Ishaq et al., 2021) authors recently employed SMOTE to balance the distribution of the data and extremely randomized trees (ET) on a subset of parameters to predict patient survival using Random Forest and attained an accuracy of 92.6%.

2.1. Proposed Methodology

The proposed model is built to get high performance for predictions and detection of disease. The following steps outlined the work.

Step 1: The data is collected from the Kaggle repository pre-processed, and normalized using a min-max scalar.

Step 2: Split the data into training and testing parts in an 80:20 ratio.

Step 3: SMOTE Synthetic Minority Oversampling technique and its variants are applied for imbalance class distribution on the training data.

Step 4: The ML Classifier is built, tested, and evaluated using an accuracy performance metric.

Step 5: Model performance is evaluated and compared.

3. MATERIAL AND METHODS

This section describes the methods, techniques, and materials(dataset) used in our work. It consists of a detailed description of the balancing techniques classification techniques and dataset description.

3.1. Balancing Techniques

Class imbalance is a major problem which leads to degrades the performance of the training model as it tends to bias towards the majority class. Class imbalance issues have been faced in various fields like Healthcare medicine, bioinformatics, and fraud detection. To address the class imbalance problem two types of approaches are suggested by the researchers. The first type is at the algorithm level the classification accuracy of the learning algorithm is increased by using cost-sensitive learning methods (Gan, Shen, An, Xu, & Liu, 2020). The second type is at the data level in this approach sampling methods are used to balance the majority and minority classes. Various oversampling methods are discussed below.

Synthetic Minority Oversampling Technique (SMOTE**)**

The synthetic Minority Oversampling technique is a pre-processing oversampling technique that deals with an imbalance class in a dataset (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). SMOTE randomly generates duplicate data of the minority class by using Euclidean distance from its nearest neighbor.

The steps involved in creating the synthetic data are given below.

Step 1: Select an instance say y randomly from the minority class: Z, $y \in Z$.

Step 2: From the current instance find the K nearest neighbors.

Step 3: Calculate the difference between one of the neighbors and the current instance and multiply it by any number gap between 0 and 1 randomly.

Step 4: A new duplicate instance is created by adding the result of step 3 to the current instance as shown in (1).

$$y\ dup = y + gap[y\ neighbor - y]$$

Here k feature, k= 1, 2,] 3,…..n of the duplicate instance.

Apart from its successful contribution to the class imbalance problem, SMOTE has some disadvantages as it oversamples every data point in the minority class with equal weightage, leading to overgeneralization, sample overlapping, the blindness of neighborhood selection, and noisy data interference. Various other oversampling techniques are discussed below which handle the problem listed above**.**

Adaptive Synthetic Sampling Approach (ADASYN) Algorithm. It is a novel adaptive synthetic sampling method to deal with imbalanced class distribution. For each minority class sample, the SMOTE algorithm generates an equal amount of synthetic data. In this approach weighted distribution is used according to the difficulty level more synthetic data is created for harder to learn and less synthetic data is created for easier ones (Khan, Xu, Khan, & Chishti, 2021). ADASYN works on density distribution and creates synthetic data in the region where the density of minority data is low and it does not create synthetic data where the density is high Apart from its successful contribution to the imbalance problem, it suffers from two major weaknesses first due to adaptability in nature its precision suffers. Second for sparsely distributed minority examples their neighborhood contains one minority data.

SMOTE-ENN (Edited Nearest neighbor)

It is a hybrid method that combines SMOTE, generates synthetic data for the minority class and ENN, and deletes certain findings from the two classes that are recognized as having a different class from the observations class and their closest neighbor majority class (Lamari et al., 2020).

The steps of SMOTE are the same as mentioned above additional steps of ENN are defined below.

(1) In A dataset with N observations, K is the number of nearest neighbors if nothing is specified then k= 3.
(2) Among all the other observations in the dataset, locate the observation's k-nearest neighbor. Then, retrieve the observations majority class from the k nearest neighbor.
(3) If the observation's k-nearest neighbor is different from the majority class then both k- the nearest neighbor and observation are recovered from the dataset.
(4) Repeat steps 2 and 3 until achieving each class's desired ratio.

3.2. Classification Techniques

This subpart describes different classification techniques used in the research work.

Logistic Regression((1)LR)

LR is a prominent supervised learning approach that determines a relationship between both independent and dependent variables by calculating the likelihood and predicting the outcome of a categorical dependent variable. The sigmoid function converts the predicted values 0, and 1 to probability.

Support Vector Machine (SVM)

SVM is a leading supervised ML algorithm used for both classification as well as regression problems. An SVM model is solely a hyperplane in multidimensional space that represents several classes. The objective function of SVM for m number of data points and the regularization coefficient which is initialized before training the SVM model and performs a tradeoff between separator margin and penalty.

K-Nearest Neighbor (KNN)

KNN is the most common Machine Learning algorithm that can be used for classification as well as regression problems. KNN is non-parametric as it does not use any data distribution assumptions on the underlying data. KNN considers the similarity between new data and existing data and puts the new data in the group that is closest to the group. In the paper (Giri, et al., 2013), KNN has been used as an automated classification technique for diagnosing coronary artery disease.

Decision Tree (DT)

A Decision Tree is a machine-learning technique used for both regressions as well as classification tasks. It resembles a tree-like structure where the root node is considered as a decision node, an internal node which is again subtree denotes dataset features branches denote decision rules and a leaf node denotes class label/outcome. In the decision tree attribute selection is done using the Attribute Selection Measure technique (Hegelich, 2017). Two metrics are used which are Information Gain and Entropy.

Random Forest (RF)

Random Forest Classifier is an ensemble technique this implies that combines multiple classifiers mainly decision trees to generate prediction results by majority voting which gives better performance than any single classifier. In the classification problem, random forest output is done by majority voting whereas in the regression problem output is calculated by the mean of the total number of the decision tree. RF is computationally expensive as it handles large datasets but simultaneously provides higher accuracy and prevents overfitting problems.

Gradient Boosting (GB)

GB is a nonlinear supervised ML algorithm. It is one of the ensemble variants where numerous weak models are created and combined to improve overall performance and higher accuracies. All the boosting algorithm works based on learning from errors of the previous model trained and trying to avoid the same mistakes made by the previous weak learning algorithm.

3.3. Dataset Description

The Heart failure prediction dataset from Kaggle is used for the experiment (Heart Failure Prediction Dataset, 2020). The dataset consists of 299 patient records with 12 features. Among them 194 are men and 105 are women. The feature death_ event is the target value where 1 indicates death and 0 indicates alive. The description of the dataset is shown in (Table 1.)

The dataset has an unequal distribution of positive class with 96 instances and negative class with 203 instances. Figure. 1. Depicts the count plot of the imbalanced dataset 0 means alive and 1 means death event due to heart disease.

Table 1. Dataset Description

Index No	Features	Summary	Valid Range
1	Age	Age in years	(40-95)
2	Anaemia	Deficiency of Haemoglobin	(0,1)
3	Creatinine Phosphokinase (CPK)	The level of CPK in the blood	(23-7861)
4	Diabetes	Diabetic or not	(0,1)
5	Ejection Fraction	Percentage of Blood Leaving	(14,80)
6	High Blood Pressure	Hypertension or not	(0,1)
7	Platelets	Platelet count	(25.01-850.00)
8	Serum Creatinine	Creatinine Level in Blood	(0.50-9.40)
9	Serum Sodium	The sodium level in the blood	(114-148)
10	Sex	Male or Female	(0,1)
11	Smoke	Patient smokes or not	(0,1)
12	Death_Event	1 indicates the patient dead 0 is alive	(0,1)

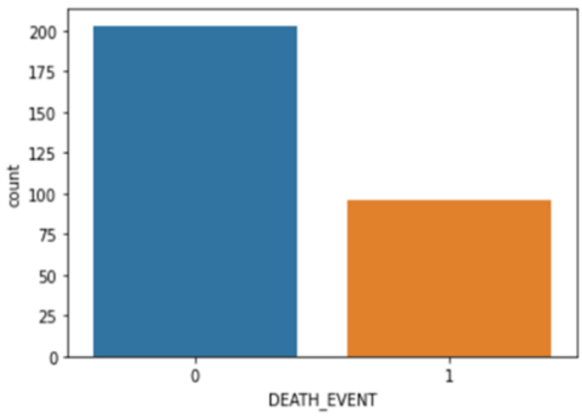

Figure 1. Count plot of the dataset

4. RESULTS AND DISCUSSION

Clearly explain the conceptual and theoretical framework, innovation description, and results: The experiment is done in Python using various libraries on a 4 GB Lenovo Intel Core i5 with 32 GB RAM. LR achieves the highest accuracy of 80% followed by SVM and KNN. Accuracy is one of the most important metrics for assessing algorithms. Experiments were done on 12 features of the heart failure dataset. Six ML algorithms namely LR, SVM, KNN, DT, RF, and GB are selected and trained on the dataset. The line graph of the accuracy metric of all classifiers is shown in Figure 2.

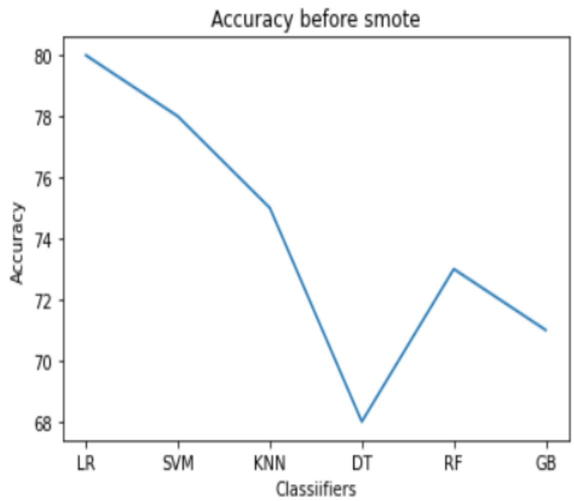

Figure 2. Line Graph of accuracy before applying the Balancing technique

Results Using SMOTE

SMOTE is an effective solution for the issue of class imbalance and has shown effective results in several areas. The line graph of the accuracy of all classifiers after applying Smote is shown in Figure 3. The accuracy of tree-based algorithms, DT, and RF increases whereas the performance of all remaining algorithms LR, SVM, KNN, and GB degrades.

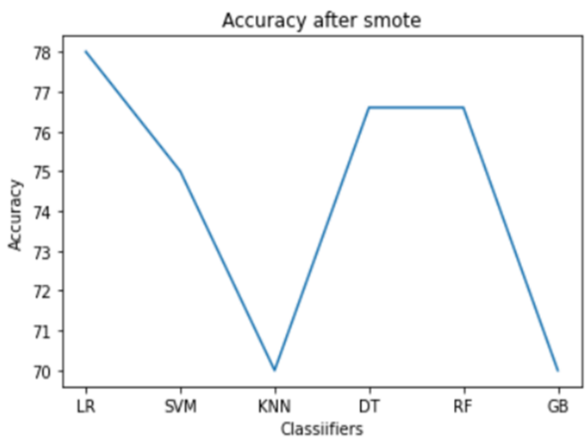

Figure 3. Accuracy after SMOTE

Results Using SMOTE_ENN

It is a hybrid technique formed by combining SMOTE and Edited nearest Neighbor (ENN). The line graph of the accuracy of all classifiers after applying SMOTE_ENN is shown in Figure 4. From the figure, it is clear that the performance of LR remains the same whereas all three tree-based algorithms increase. Accuracies of SVM and KNN degrades.

Results Using ADASYN

It is an adaptive synthetic sampling approach used to solve imbalanced class problems. A line graph of the accuracy of all classifiers after applying ADASYN is shown in Figure 5. KNN, DT, RF, and GB perform better with the ADASYN technique.LR does not show any improvement. Compared to other techniques in the work ADASYN performs better for the given dataset.

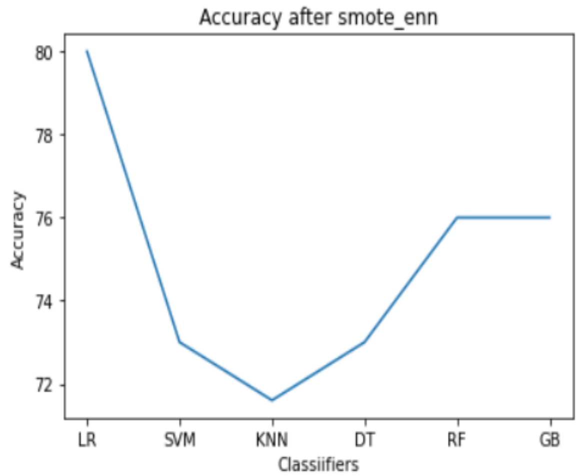

Figure 4. Line graph of accuracy after SMOTE_ENN

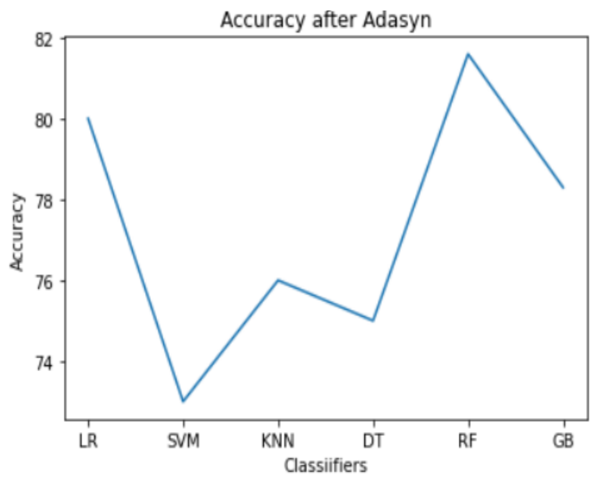

Figure 5. Line graph of accuracy after ADASYN

Comparison in the form of multiple bar graph between all the techniques discussed in the paper namely SMOTE, SMOTE_ENN, and ADASYN is shown in Figure 6, whereas Table.2. gives the summarized value corresponding to Figure 6.

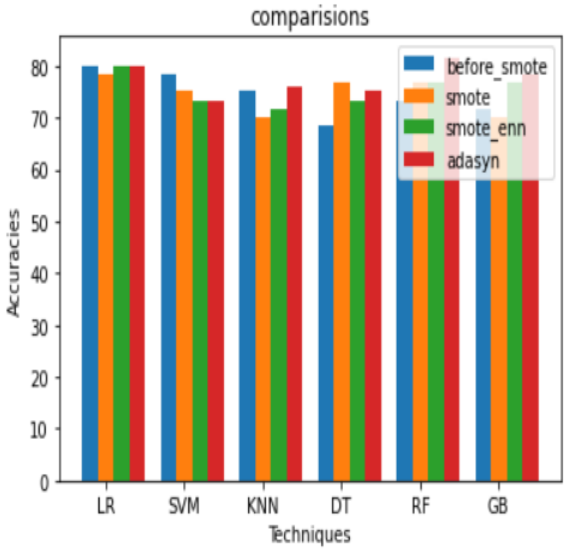

Figure 6. Comparison of all techniques

Table 2. Comparison of all Techniques

S.No	Classifier	Accuracy before Balancing techniques	Accuracy after applying SMOTE	Accuracy after applying SMOTE-ENN	Accuracy after ADAsyn
1	LR	.80	.78	.80	.80
2	SVM	.78	.75	.73	.73
3	KNN	.75	.70	.71	.76
4	DT	.68	.76	.76	.75
5	RF	.73	.76	..76	.81
6	GB	.71	.70	.76	.78

5. CONCLUSION AND FUTURE WORK

Heart disease has grown into one of the world's leading causes of death and a looming threat to global health. In this crucial condition ML model can be very helpful in predicting diseases at their earliest phases. This paper incorporates several kinds of ML classification strategies and balancing techniques to identify and foresee heart disease. The Heart Failure prediction dataset is used for experimentation. SMOTE, SMOTE-ENN, and ADASYN balancing techniques are applied on six state-of-the-art algorithms, the performance of the techniques are compared and results show that ADASYN techniques perform better in four classifiers namely KNN, DT, RF, and GB for the given dataset.

In the future, multiclass classification of some real-world datasets related to cardiac disease can be used for research. Additionally, further research in comparison and analysis using various hyperparameter optimization techniques and outlier detection methods might be done. In addition, we will explore more ML and hybrid classifiers with other oversampling and under-sampling methods to increase the predictive model's performance. Besides, Blockchain technology (Fatima, Agarwal, & Sohail, 2022) for effective healthcare can be explored as well by researchers to solve several issues that grip the healthcare industry.

Acknowledgement:
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper."
"This research did not receive any grants from governmental, private, or nonprofit funding bodies."

References:

Abdellatif, A., Abdellatef, H., Kanesan, J., Chow, C. O., Chuah, J. H., & Gheni, H. M. (2022). An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods. *IEEE Access, 10*, 79974-79985. doi:10.1109/ACCESS.2022.3191669

Ashri, S. E., EL-Gayar, M. m., & El-Daydamony, E. M. (2021). HDPF: Heart Disease Prediction Framework Based on Hybrid Classifiers and Genetic Algorithms. *IEEE Access, 9*, 146797-146809. doi: 10.1109/ACCESS.2021.3122789

Cardiovascular diseases. (n.d.). Retrieved from WHO: https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). [Submitted on 9 Jun 2011]. *Journal Of Artificial Intelligence Research, 16*, 321-357. doi:10.1613/jair.953

Dhand, G., Sheoran, K., Agarwal, P., & Biswas, S. S. (2022). Deep enriched salp swarm optimization based bidirectional -long short term memory model for healthcare monitoring system in big data. *Informatics in Medicine Unlocked, 32*, 101010. doi:10.1016/j.imu.2022.101010

Fang, J., Luncheon, C., Ayala, C., Odom, E., & Loustalot, F. (2019). Awareness of Heart Attack Symptoms and Response Among Adults - United States, 2008, 2014, and 2017. *MMWR Morb Mortal Wkly Rep, 68*(5), 101-106. doi:10.15585/mmwr.mm6805a2

Fatima, N., Agarwal, P., & Sohail, S. S. (2022). Home ICT Analysis and Applications. In S. Fong, N. Dey, & A. Joshi, *ICt Analysis and Application* (pp. 193-201). Singapore: Springer.

Fitriyani, N. L., Syafrudin, M., Alfian, G., & Rhee, J. (2020). HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System. *IEEE Access, 8*, 1333034-133050. doi:10.1109/ACCESS.2020.3010511

Gan, D., Shen, J., An, B., Xu, M., & Liu, N. (2020). Integrating TANBN with cost sensitive classification algorithm for imbalanced data in medical diagnosis. *Computers & Industrial Engineering, 140*, 106266. doi:10.1016/j.cie.2019.106266

Garate-Escamila, A. K., Hassani, A. H., & Andres, E. (2020). Classification models for heart disease prediction using feature selection and PCA. *Informatics in Medicine Unlocked, 19*, 100330. doi:10.1016/j.imu.2020.100330

Giri, D., Acharya, U. R., Martis, R. J., Sree, S. V., Lim, T.-C., Ahamad vI, T., & Suri, J. S. (2013). Automated diagnosis of Coronary Artery Disease affected patients using LDA, PCA, ICA and Discrete Wavelet Transform. *Knowledge-Based Systems, 37*, 274-282. doi:10.1016/j.knosys.2012.08.011

Gupta, A., Kumar, R., Arora, H. S., & Raman, B. (2020). MIFH: A Machine Intelligence Framework for Heart Disease Diagnosis. *IEEE Access, 8*, 14659-14674. doi:10.1109/ACCESS.2019.2962755

Guyon, I., & Elisseef, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157-1182.

Heart Failure Prediction Dataset. (2020, January 5). Retrieved from Kaggle: https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data

Hegelich, S. (2017). Decision Trees and Random Forests: Machine Learning Techniques to Classify Rare Events. *European policy analysis*, 98-120. doi:10.18278/epa.2.1.7

Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques. *9*, 39707-39716. doi:10.1109/ACCESS.2021.3064084

Khadir, A. S., & Amanullah, K. M. (2018). A PSO-SVM FEATURE SELECTION METHOD FOR IMPROVING THE PREDICTION ACCURACY OF HEART DISEASE. *International Journal of Computer Engineering and Technology (IJCET), 9*(4), 96-104. doi:10.34218/IJCET.9.4.2018.011

Khan, T. M., Xu, S., Khan, Z. G., & Chishti, M. U. (2021). Implementing Multilabeling, ADASYN, and ReliefF Techniques for Classification of Breast Cancer Diagnostic through Machine Learning: Efficient Computer-Aided Diagnostic System. *J Healthc Eng*, 5577636. doi:10.1155/2021/5577636.

Lamari, M., Azizi, N., Hammami, N. E., Boukhamia, A., Cheriguene, S., dendani, N., & Benzebouchi, N. E. (2020). SMOTE-ENN Based Data Sampling and Improved Dynamic Ensemble Selection for Imbalanced Medical Data Classification. *Advances on Smart and Soft Computing Proceedings of ICACIn 2020* (pp. 37-49). casablanca Morocco: Proceedings of ICAI.

Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques. *IEEE Access, 7*, 81542-81554. doi: 10.1109/ACCESS.2019.2923707

Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Machine Learning Technology-Based Heart Disease Detection Models. *J Healthc Engg, 2022*, 2022. doi:10.1155/2022/7351061

Phegade, K. S., Gurule, D. D., Jadhav, D. G., Jadhav, R. S., Patil, P. G., & Thorat, D. Y. (2019). Heart Attack Prediction System using Artificial Neural Network. *International Journal for Research in Applied Science & Engineering Technology (IJRASET), 7*(2), 971-974. doi:10.22214/ijraset.2019.2151

Ramana, K., Kumar, M. R., Sreenivasulu, K., Gadekallu, T. R., Bhatia, S., Agarwal, P., & Idress, S. m. (2022). Early Prediction of Lung Cancers Using Deep Saliency Capsule and Pre-Trained Deep Learning Frameworks. *Front Oncol, 12*, 886739. doi:10.3389/fonc.2022.886739

Rani, P., Kumar, R., Ahmed, N. M., & Jain, A. (2021). A decision support system for heart disease prediction based upon machine learning. *Journal of Reliable Intelligent Environments, 7*(3), 263-275. doi:10.1007/s40860-021-00133-6

Shah, S. M., Batool, S., Khan, I., Ashraf, M. U., Abbas, S. H., & Hussain, S. A. (2017). Feature extraction through parallel Probabilistic Principal Component Analysis for heart disease diagnosis. *Physica A: Statistical Mechanics and its Applications, 482*, 796-807. doi:10.1016/j.physa.2017.04.113

Tama, B. A., Im, S., & Lee, S. (2020). Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble. *Bio Med Research International*, 9816142-9816152. doi:10.1155/2020/9816142

Uzama Sadar
Jamia Hamdard,
New Delhi,
India
life.uzma@gmail.com
ORCID 0009-0008-9959-756X

Parul Agarwal
Jamia Hamdard,
New Delhi,
India
pagarwal@jamiahamdard.ac.in
ORCID 0000-0002-7297-335X

Suraiya Parveen
Jamia Hamdard,
New Delhi,
India
suraiya@jamiahamdard.ac.in
ORCID 0000-0001-5499-6999

Geetika Dhand
Maharaja Surajmal Institute of Technology,
New Delhi,
India
geetika.dhand@msit.in
ORCID 0000-0002-7752-2515

Kavita Sheoran
Maharaja Surajmal Institute of Technology,
New Delhi,
India
Kavita.sheoran@msit.in
ORCID 0000-0001-7985-9288