Medical Diagnosis Support System for Cardiovascular Disease Prediction Machine Learning Based

Received: 1 January 2023; Accepted: 6 March 2023 *Research Article*

Nidhal Mohsin Hazzaa
Department of Computer Science
College of Computer Science and Information Technology
Kirkuk University ; Gazi University
Kirkuk, Iraq ; *Ankara*, Türkiye
nmohsin.hazzaa@gazi.edu.tr
0000-0001-7071-922X

Oktay Yıldız
Department of Computer Engineering
Gazi University
Ankara, Türkiye
oyildiz@gazi.edu.tr
0000-0001-9155-7426

***Abstract*— Early prediction and diagnosis of CVD are crucial for the effective management and prevention of advanced cases. In this study, a diagnosis system using supervised machine learning is proposed to predict CVD. The system employs multiple ML classifiers, including RF, DT, SVM, LR, and MLP, for predicting atherosclerosis. The UCI repository Sani Z-Alizadeh dataset was used for this research. The imbalanced nature of the dataset, which refers to the number of instances belonging to one class being significantly greater than the number of instances belonging to another class, was addressed using the Synthetic Minority Oversampling Technique (SMOTE) for data resampling. Ten-fold cross-validation procedures were used to split the dataset. The performance of the five machine learning (ML) classifiers was evaluated using standard performance metrics. The evaluation revealed that all classifiers achieved a performance improvement of at least 2%. The proposed model has potential applications in healthcare and can improve clinical diagnosis of CVD disorders, leading to optimized diagnosis, prevention of advanced cases, and lower treatment expenses.**

Keywords*— heart disease, medical diagnosis support system (MDSS), clinical data, machine learning

I. INTRODUCTION

The heart serves as a vital component in maintaining the proper function of the human body, as it facilitates the circulation of oxygenated blood through the arteries and veins to all body tissues. Any disorder that disrupts the heart's ability to pump blood effectively is generally referred to as heart disease [1]. Sadly, heart disease remains a significant contributor to global mortality rates, with an alarming 17.9 million individuals succumbing to this condition annually, as reported by the World Health Organization in 2021 [2]. Heart disease manifests in various forms, including but not limited to coronary artery disease, congenital heart disease, arrhythmia, and myocardial infarction, each presenting unique challenges to diagnosis and treatment. Heart disease is a complex ailment influenced by various risk factors, which can be categorized as behavioral, genetic, and physiological. Behavioral risk factors such as smoking, excessive alcohol and caffeine consumption, stress, and physical inactivity can contribute to the development of heart disease. Genetic factors can also predispose individuals to heart disease. Physiological variables, including but not limited to obesity, hypertension, and high-cholesterol. The patient with cardiac has several symptoms such as chest pain, dizzy sensations, and deep sweating [3]. Diagnosing at an early stage can reduce the number of deaths. Taking preventive actions is made possible in large part by the ability to accurately and quickly diagnose cardiac disease. Especially, in developing nations, there is a shortage of medical professionals and proper medical centers in remote areas. Due to the numerous limitations of manual detection of CVD, scientists have shifted their focus to new technologies such as Data Mining, Machine Learning, and Deep Learning to automate disease classification and prediction [4]. Automation combined with ML and DL can be used to develop a support system that can quickly and cost-effectively detect a cardiac disease from clinical data. These have proven to be useful in assisting decision-making and forecasting from the massive amounts of data generated by the healthcare business [5]. The identification of atherosclerosis risk factors is based on medical experts' and doctors' knowledge and expertise, and these risk factors are classified as either controllable or uncontrollable. Several characteristics are utilized to identify these factors, with family history, age, and gender being unmodifiable risk factors for atherosclerosis [6].

The structure of this paper is as follows: In Section II, related works in the literature are reviewed. Section III presents and explains the methodology of the proposed system, including the selected machine learning algorithms and the evaluation parameters used to estimate and compare the performance of the proposed MDSS with similar measures. Section IV describes the CAD datasets used, implementation details, and the results are discussed. Finally, Section V concludes this research and provides future perspectives for further research.

II. RELATED WORKS

The classification and prediction of heart disease diagnosis has been the focus of numerous studies employing various ML models. Ali et al. [7] used the KNN, RF, and DT classifiers to produce top-notch outcomes. In addition, for all algorithms other than MLP and KNN, feature importance ratings were estimated, and these features were sorted according to their significance scores. Pavithra et al. [8] proposed a novel hybrid feature selection strategy, named HRFLC, which merges RF, AdaBoost, and utilized filter, wrapper, and embedding approaches to select eleven features, which resulted in a 2% increase in hybrid model accuracy. Kolukisa et al. [9] have presented six classifiers, FS method, and a probabilistic (FS) approach. After hyperparameter adjustment, Saboor et al. [10] applied nine machine learning classifiers to the final dataset. They applied standard K-fold cross-validation methods to confirm their findings. With hyperparameter adjustment and

data normalization, the accuracy improved dramatically. Türkmenoğlu et al. [11] suggested a Heart failure survival analysis utilizing the Correlation Matrix and RF techniques. Due to the uneven class distribution of the data set, data cleansing, oversampling, and undersampling were applied. They demonstrated that removing the class imbalance from the data set improved the performance of the classifiers.. In [12], the authors presented a novel, optimized algorithm which employed many classifier techniques, such as NB, KNN, Bayesian Optimized (BO-SVM), and (SSA-NN). The results indicated that the BO-SVM classifier performed the best with an accuracy of 93.3%, followed by the SSA-NN classifier with an accuracy of 86.7%. Sudha and colleagues [13] introduced a hybrid machine learning system that integrates (CNNs) and (LSTM) to improve the accuracy of classification on datasets. The researchers validated the performance of this hybrid model using the k-fold cross-validation method with 89% of an accuracy rate. The authors of [14] described a ML strategy by using SVM, NB, and DT algorithms, they emphasized the use of polynomial regression in predicting vital signs, taking into consideration the nonlinear character of these variables. Perumal et al. [15] developed CVD dataset to improve model performance and feature quality, the authors suggested feature standardization, PCA-based feature reduction, and entropy-based feature engineering (FE). They trained ML classifiers using seven main components. The study found that LR and SVM classifiers were virtually as accurate as KNN. Research conducted in [16], the authors proposed imputing missing values (IMV) and removing outliers (OR). Experimental findings indicated that the suggested model (LR + NB) outperformed high results in all metrics, particularly in terms of AUC and accuracy. A comprehensive study examined how numerical, categorical, and combination numerical and categorical feature types affect machine learning algorithms [17]. Gradient Boosting, AdaBoost, CatBoost, XGBoost, ANN, RF, SVM, DC, and LR classifiers were compared. The study also indicated that SVM and AdaBoost ensemble learning with categorical features performed best for CVD prediction. Table 1 summarizes relevant empirical research studies on heart disease prediction.

III. Theoretical Background

This section presents an overview of the techniques and tools utilized in our experiment, which aimed to detect cardiac disease using machine learning models. First, we describe the machine learning models employed in the experiment. We then outline the evaluation metrics used to measure the performance of the models.

A. Machine Learning Classifiers

Various machine learning algorithms have developed over time for heart disease diagnoses. Most researchers used more than one ML classifier in their papers to select the accurate one. The five classification techniques utilized in this research including their specific features and parameters are as followed:

1) Random Forest(RF)

Is a decision-tree based ML model. The technique randomly selects training papers from the feature space's m-try dimension subspace and calculates every probability using m-try features. Leaf nodes divide data best until saturation. An ensemble of K unpruned trees h1(X1), h2(X2),... hk(Xk) yields the greatest likelihood of classification, making RF a powerful classification method for textual data with many dimensions. [18].

2) Decision Tree(DT)

It is a tree-like model that classifies data points based on their node requirements [19]. As information passes through the DT's internal nodes, it gets categorized. For the dividing criterion, the Gini index [1,2] is used. Gini indices are determined per attribute. The least Gini Coefficient attribute would partition the data [20]. A tree is formed by repeatedly selecting the lowest Gain ratio characteristic.

3) Logistic Regression(LR)

Is a widely used statistical method for binary classification problems. LR uses a logistic function to restrict the output of a linear equation to the range of 0 and 1. The key difference between linear and probabilistic regression is that LR is limited to a binary (0 or 1) spectrum. The exponentiated LR slope coefficient (eb) can be easily interpreted as an odds ratio as mentioned in Eq.1, which is a significant advantage of LR over other methods such as probit regression. [21].

$$Logistic\ Function = \frac{1}{1+e^{-x}} \quad (1)$$

4) Support Vector Machine (SVM)

Is a prominent kernel-based learning technique for image classification and other ML tasks. SVMs solve a convex quadratic optimization problem to find a globally optimal solution [22]. SVM assumes prior knowledge of the data distribution and creates a hyper-plane with a maximally broad margin to classify data into distinct categories or keep similar data of one kind on one side and similar data of another type on the other. [[23], a linear SVM can be described by the following Eq.2:

$$f(x) = sign(w^T x + b) \quad (2)$$

TABLE I. Comparative review of recent research studies and clinical guidelines for HD diagnosis.

Ref.	Year	Technique(s)	Dataset	Accuracy
[7]	2021	LR , ABM, MLP, KNN, DT, RF	Hungary, Switzerland, Cleveland, and Long Beach.	97.08%
[8]	2021	RF, AD, PC(PEARSON COEFFICIENT)	UCI Repository	81%
[9]	2023	SVM, MLP, RF, KNN, LR, LDA	Z-Alizadeh Sani, cleveland, Statlog	87.6%
[10]	2022	LR, ET, MNB, CART, SVM, LDA, AB, RF, XGB	Z-lizadeh Sani, Statlog	91.50%
[11]	2021	RF, KNN, ET	faisalabad cardiology hospital	84.58%
[12]	2021	NB, BO-SVM, KNN, SSA-NN	UCI Repository	93.3%
[13]	2023	CNN , LSTM	Cleveland, Hungar	89%
[14]	2021	SVM, NB, DC (J48)	Universityof Queensland	-
[15]	2020	LR, SVM, KNN	Cleveland	80.33%
[16]	2022	(LR+NB)	CHDD, HHDD, SHDD and VAMC	92.7%
[17]	2022	GB,XGBoost,LR AdaBoost, CatBoost, MLP, RF, SVM, DC,	Cleveland	82%

5) Multilayer perceptron (MLP)

MLP is supervised using hidden synthetic neuron layers. Perceptrons stimulate each neuron. Neuron-like perceptrons. The activation function assigns weighted inputs to two levels per neuron. Weight changes teach perceptrons [24]. MLPs use past data to generate output outcomes when the desired outcome is ambiguous. Data must match input and output values.

B. *Evaluation metrics*

Evaluation metrics are measures that are used to evaluate the performance of a machine learning model. These metrics provide a quantitative way to assess how well the model is performing on the given task, such as classification or regression. Some common evaluation metrics include:

- Accuracy: It is the percentage of correctly predicted labels among all the predictions as mentioned in Eq.3.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

- *Precision:* The percentage of true positive predictions among all the positive predictions. Precision measures the model's ability to correctly identify positive cases Eq. 4 [25].

$$\text{Precision}=\frac{TP}{TP+FP} \quad (4)$$

- *Sensitivity:* Also, called Recall, The percentage of true positive predictions among all the actual positive cases. Recall measures the model's ability to identify all positive cases, as illustrated in Eq. 5 [26].

$$\text{Sensitivity}=\frac{TP}{TP+FN} \quad (5)$$

- *F1-score:* The harmonic mean of precision and recall. It provides a single metric that balances precision and recall, and evaluates the classification model's performance in the imbalanced classes, as shown in Eq.6 [27].

$$\text{F1-score}= 2\times \frac{\text{precision}\times\text{sensitivity}}{\text{precision}+\text{sensitivity}} \quad (6)$$

- *(MCC):* Matthew's Correlation coefficient which provides a balanced measure of the model's performance across both positive and negative classes and evaluates the quality of a binary classifier in case of imbalanced classes [28], as presented in Eq.7.

$$\text{MCC} =\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TP+FP)(TN+FN)}} \quad (7)$$

IV. Experimental Study And Result

The experiments were conducted through the WEKA tool (Waikato Environment for Knowledge Analysis), an open-source JAVA-based software capable of applying algorithms directly to a dataset or via JAVA code for data pre-processing. To split the dataset into a training set and a test set, 10-fold cross-validation was employed. Furthermore, this section includes details on the datasets used, data pre-processing techniques applied, and the analysis of findings using the proposed framework. The algorithmic operations of the proposed model are provided in Algorithm 1.

Algorithm 1 Proposed support system for heart disease diagnosis
Input: Sani Z-Alizadeh dataset
Begin

1. Data pre-processing:
   a. resampling imbalanced data using(SMOT)
   b. deploy data normalization
2. Split dataset by 10-cross validation
3. Classification model:
   a. perform a ML classifier
   b. log the classifier performance
   c. repeat a-b until all classifier are deployed
4. Performance measured using five metrics
   (Accuracy, Recall, Precision, F1-score and Mcc)

End.

A. *Dataset*

This study made use of the UCI repository's Z-Alizadeh Sani dataset on heart disease, which includes 216 patients with heart disease and 87 healthy individuals. The dataset contains 54 clinical and demographic features, which are divided into 23 numerical and 31 categorical data. Table 2 provides a comprehensive list of these features and explains in detail the characteristics selected for the study [29]. As illustrated in Fig. 1 the pie chart represents the gender distribution of the cases in the targeted dataset. The data reveals that males comprise 58% of the cases, while females make up 42%, indicating a significant gender imbalance in the dataset.

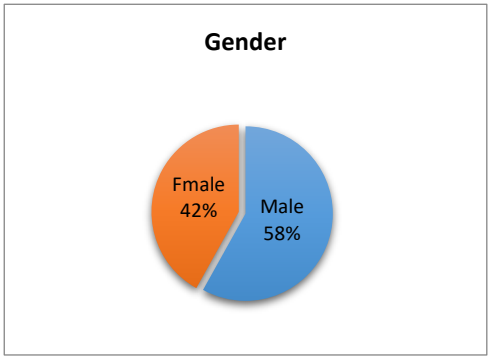

Fig. 1. gender distribution within dataset

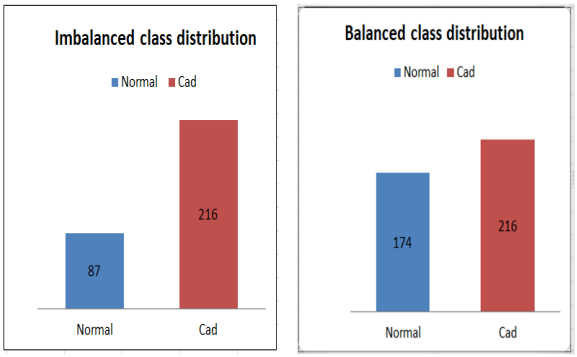

Fig. 2. Class distribution of Normal and Cad

TABLE II. FEATURES OF Z-ALIZADEH SANI DATASET

Feature type	Feature name	Range
Demographic	Age	30-86
	Weight	48-120
	Length	140-188
	Sex	Male,Female
	BMI(Body Mass Index)	18.1-40.9
	DM (Diabetes Mellitus)	Yes,No
	HTN (Hyper Tension)	Yes,No
	Current Smoker	Yes,No
	Ex-Smoker	Yes,No
	FH (Family History)	Yes,No
	Obesity	Yes,No
	CRF (Chronic Renal Failure)	Yes,No
	CVA (Cerebrovascular Accident)	Yes,No
	Airway Disease	Yes,No
	Thyroid Disease	Yes,No
	CHF (Congestive Heart Failure)	Yes,No
	DLP (Dyslipidemia)	Yes,No
Clinical	BP (Blood Pressure: mmHg)	90-190
	PR (Pulse Rate) (ppm)	50-110
	Edema	Yes,No
	Weak peripheral pulse	Yes,No
	Lung Rales	Yes,No
	Systolic murmur	Yes,No
	Diastolic murmur	Yes,No
	Typical Chest Pain	Yes,No
	Dyspnea	Yes,No
	Function Class	1,2,3,4
	Atypical	Yes,No
	Nonanginal	Yes,No
	Exertional CP (Exertional Chest Pain)	Yes,No
	LowTH Ang (low Threshold angina)	Yes,No
	Rhythm	Yes,No

B. Data pre-processing

In order to address the notable uneven distribution of classes within the dataset, we specify that the (Normal) category pertains to patients who do not have any cardiovascular disease, while the (Cad) category refers to those who do, as illustrated in Fig. 2(a). It should be emphasized that the dataset contains about three times more individuals with CVD than those without it. At this phase, the challenge of class imbalance is resolved by employing the synthetic minority oversampling method (SMOTE). It is an oversampling technique that has gained considerable use in the medical domain for handling imbalanced data [30]. By producing minority class random synthetic data from its closest neighbors using Euclidean distance, SMOTE augments the quantity of data instances. New instances begin to resemble the original data since they are formed based on the original data [31]. A fresh training dataset is created in this work utilizing the SMOTE approach. Each class's data sample size was increased by SMOTE from 303 to 390 as shown in Fig.2 (b). Then, 10-fold cross-validations have been performed to split the dataset into test and train sets.

V. RESULT AND DISCUSSION

In order to evaluate the effectiveness of the five proposed classifiers for diagnosing cardiac illness, different metrics such as specificity, precision, recall, F1-score, Mcc, and accuracy were utilized. Moreover, we compared the model's results before and after attempted to strike a balance in the dataset.

The findings revealed that certain algorithms demonstrated strong accuracy, while others performed poorly prior to balancing the data through the use of over-sampling. Table 3 displays the performance of the utilized classifiers on the raw data (imbalanced data), demonstrating that SVM had the highest accuracy performance at 86.798% and other metrics.

Table 4 presents the performance of the classifiers on balanced data. A noticeable improvement in the performance of all classifiers across all metrics was observed with 10-fold cross-validation. For instance, RF's accuracy improved from 85% to 90%, DT's accuracy improved from 79% to 84%, LR's accuracy increased from 83% to 86%, SVM's accuracy improved from 86.79% to 87.69%, and MLP's accuracy increased from 82% to 85%.

It is noteworthy that improvements in all evaluation metrics, particularly in MCCs performance, were achieved. Table 5 and Fig. 3 provide a comparison of the accuracy of the ML classifiers before and after balancing the dataset.

Finally, to provide a more comprehensive comparison with previous studies, we discuss studies that have used the imbalanced dataset and the same resampling (SMOT) techniques [9, 11] from Table 1, it is evident that our proposed MDSS has a higher accuracy with 90.51% over algorithms of [9] with 87.6% accuracy. Moreover, our approach outweighed the study [11], although they used more than resampling techniques over the dataset and three classifiers.

TABLE III. ML ALGORITHMS PERFORMANCE MEASURES WITH TEST MODEL 10-FOLD CROSS-VALIDATION (IMBALANCED DATA)

Classifier	Accuracy	Precision	Recall	F-Measure	MCC
SVM	86.798 %	0.911	0.903	0.907	0.680
RF	85.808%	0.865	0.949	0.905	0.637
DT	79.207 %	0.837	0.880	0.858	0.474
LR	83.168 %	0.884	0.880	0.882	0.590
MLP	82.178 %	0.879	0.870	0.874	0.568

TABLE IV. ML ALGORITHMS PERFORMANCE MEASURES WITH TEST MODEL 10-FOLD CROSS-VALIDATION (BALANCED DATA)

Classifier	Accuracy	Precision	Recall	F-Measure	MCC
SVM	87.692 %	0.893	0.884	0.888	0.751
RF	90.512%	0.909	0.921	0.915	0.808
DT	84.615%	0.861	0.861	0.861	0.689
LR	86.666 %	0.887	0.870	0.879	0.731
MLP	85.641 %	0.908	0.824	0.864	0.716

TABLE V. COMPARISON IN ACCURACY OF CLASSIFIERS WITH IMBALANCED AND BALANCING

Classifier	Accuracy with 10-fold cross-validation	
	Imbalanced dataset	Balanced dataset
SVM	86.798 %	87.692 %
RF	85.808%	90.512%
DT	79.207 %	84.615%
LR	83.168 %	86.666 %
MLP	82.178 %	85.641 %

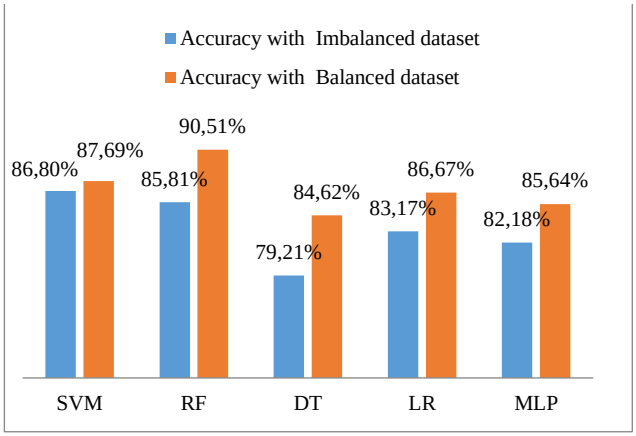

Fig. 3. Accuracy of classifiers with 10-fold cross-validation for imbalanced and balanced dataset

VI. Conclusion & Future Work

The progress in ML techniques has made it possible to use data mining effectively in healthcare. The study conducted aimed to develop a model for diagnosing heart disease using biochemical values at a lower cost. The results showed that all classifiers had higher accuracy rates when subjected to a 10-fold cross-validation test model after balancing the dataset. The study highlights the importance of addressing the class imbalance in preparing datasets for effective machine learning and presents a methodology for using multiple classifiers to predict CVD. This proposed methodology has the potential to enhance diagnostic accuracy, detect patients at an early stage, decrease mortality rates, and enable further treatment, especially in situations with imbalanced datasets. As technology advances, future studies should aim to expand this approach to larger datasets and leverage deep learning principles. This will allow for even more accurate diagnoses, better patient outcomes, an overall improvement in healthcare services, and an improved quality of life for people worldwide.

References

[1] 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.

[2] https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) ,11 June 2021.

[3] Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. SN Computer Science, 1(6), 1-6.

[4] Swathy, M., & Saruladha, K. (2021). A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques. ICT Express.

[5] Baghel, N., Dutta, M. K., & Burget, R. (2020). Automatic diagnosis of multiple cardiac diseases from PCG signals using convolutional neural network. Computer Methods and Programs in Biomedicine, 197, 105750.

[6] Nangia, R., Singh, H., & Kaur, K. (2016). Prevalence of cardiovascular disease (CVD) risk factors. medical journal armed forces india, 72(4), 315-319.

[7] Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M., & Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. Computers in Biology and Medicine, 136, 104672.

[8] Pavithra, V., & Jayalakshmi, V. (2021). Hybrid feature selection technique for prediction of cardiovascular diseases. Materials Today: Proceedings.

[9] Kolukisa, B., & Bakir-Gungor, B. (2023). Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis. Computer Standards & Interfaces, 84, 103706.

[10] Saboor, A., Usman, M., Ali, S., Samad, A., Abrar, M. F., & Ullah, N. (2022). A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms. Mobile Information Systems, 2022.

[11] Türkmenoğlu, B. K., & Yildiz, O. (2021, June). Predicting the survival of heart failure patients in unbalanced data sets. In 2021 29th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.

[12] Patro, S. P., Nayak, G. S., & Padhy, N. (2021). Heart disease prediction by using novel optimization algorithm: A supervised learning prospective. Informatics in Medicine Unlocked, 26, 100696.

[13] Sudha, V. K., & Kumar, D. (2023). Hybrid CNN and LSTM Network For Heart Disease Prediction. SN Computer Science, 4(2), 172.

[14] Shah, W., Aleem, M., Iqbal, M. A., Islam, M. A., Ahmed, U., Srivastava, G., & Lin, J. C. W. (2021). A Machine-Learning-Based System for Prediction of Cardiovascular and Chronic Respiratory Diseases. Journal of Healthcare Engineering, 2021.

[15] Perumal, R., & Kaladevi, A. C. (2020). Early prediction of coronary heart disease from cleveland dataset using machine learning techniques. Int. J. Adv. Sci. Technol, 29, 4225-4234.

[16] Rajendran, R., & Karthi, A. (2022). Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers. Expert Systems with Applications, 207, 117882.

[17] Pan, C., Poddar, A., Mukherjee, R., & Ray, A. K. (2022). Impact of categorical and numerical features in ensemble machine learning frameworks for heart disease prediction. Biomedical Signal Processing and Control, 76, 103666.

[18] Jackins, V., Vimal, S., Kaliappan, M., & Lee, M. Y. (2021). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. The Journal of Supercomputing, 77(5), 5198-5219.

[19] Chinnasamy, P., Kumar, S. A., Navya, V., Priya, K. L., & Boddu, S. S. (2022). Machine learning based cardiovascular disease prediction. Materials Today: Proceedings.

[20] Gao, C., & Elzarka, H. (2021). The use of decision tree based predictive models for improving the culvert inspection process. Advanced Engineering Informatics, 47, 101203.

[21] Schober, P., & Vetter, T. R. (2021). Logistic regression in medical research. Anesthesia and analgesia, 132(2), 365.

[22] Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408, 189-215.

[23] Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 6308-6325.

[24] Valupadasu, R., & Chunduri, B. R. R. (2019, May). Automatic classification of cardiac disorders using MLP algorithm. In 2019 Prognostics and System Health Management Conference (PHM-Paris) (pp. 253-257). IEEE.

[25] Juba, B., & Le, H. S. (2019, July). Precision-recall versus accuracy and the role of large data sets. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 4039-4048).

[26] Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.

[27] Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC genomics, 21(1), 1-13.

[28] Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData mining, 14(1), 1-22.

[29] https://archive.ics.uci.edu/ml/datasets/Z-Alizadeh+Sani.

[30] Blagus, R., & Lusa, L. (2015). Joint use of over-and under-sampling techniques and cross-validation for the development and assessment of prediction models. BMC bioinformatics, 16(1), 1-10.

[31] Kovács, G. (2019). An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced datasets. Applied Soft Computing, 83, 105662