**An empirical analysis of ML techniques and/or algorithms for disease diagnosis prediction from the perspective of cardiovascular disease (CVD)**

**Anil Kumar Prajapati1 and 2Umesh Kumar Singh2**

1[anilprajapatiujjain@gmail.com](mailto:anilprajapatiujjain@gmail.com), 2[umesh.icsvv@gmail.com](mailto:umesh.icsvv@gmail.com)

*2Institute of Computer Science, Vikram University Ujjain, MP, India.*

***Abstract: In recent times the death rate from heart disease (i.e., cardiovascular) is abnormally high. At present, heart disease has emerged as a serious disease for mankind, and prevention of which is very important in time. Diseases prediction through machine learning and/or deep learning are increasingly popular in recent years and there is a number of major ongoing research on predicting heart disease prediction. Predicting heart disease and/or cardiovascular through machine learning is a low costly and time-efficient method. ML techniques are based on the principle that computers recognize data and assign tasks automatically and/or with human input. Machine learning is a mathematical framework that combines mathematical, statical, and optimization techniques to predict outcomes based on input data (features, attributes, factors). As machine learning techniques have evolved, human diseases can now be detected more accurately and efficiently. A comprehensive review of various methods for predicting cardiovascular disease using machine learning is provided in this paper.***

***Index Terms- CVD, Machine learning, Health Care, Ensemble Technique, ML Algorithms, Confusion Matrix.***

1. **INTRODUCTION**

Heart disease is a serious threat to mankind and it's a big challenge and concern for researchers in the current scenario. The high mortality rate due to cardiovascular disease remains a major concern worldwide. High blood pressure, diabetes, cholesterol, chest pain, obesity, genetics, etc. are seen as the main risk factors for cardiovascular disease [1-2]. According to the report published by the World Health Organization (WHO) on June 11, 2021, about 17.9 million deaths occurred due to cardiovascular disease in the year 2019 which accounts for approximately 32% of all deaths globally [3]. Also, according to the study released by Global Burden of Disease, India has the highest number of age-standardized cardiovascular disease (CVD) deaths per 100,000 population at 272, while the global number of deaths is 235 per 100,000 population. The mortality rate due to cardiovascular disease (CVD) is on the rise in India [4]. A. Srinivas et al. in their report pointed out that various diseases such as cardiovascular disease, respiratory diseases, cancer, cerebrovascular diseases, stroke, and diabetes cause the death of 17.7 million people worldwide. About a fifth of this occurs in India, especially among the youth. various risk elements contribute to the increased prevalence of CVD in Indians, such as blood pressure, diabetes, dyslipidemia, smoking, and fatness [5]. The American Heart Association has released a statistics fact sheet on death from cardiovascular disease. This fact shows the causes of global death due to heart diseases such as coronary cardiovascular disease (CHD), stroke, sudden cardiac arrest, heart disease, stroke, and heart disease risk factors, nutrition, overweight/obesity, cholesterol, Diabetes, blood pressure (HBP), smoking, physical inactivity, and genetic causes. In this report, the association provides a detailed analysis of the total CVD deaths between 2015 and 2021[6]. Various reports published on the risk of heart disease clarify the severity of heart disease. ML techniques are increasingly used in medical diagnosis and disease prediction which is a field of artificial intelligence. In recent years, advances in computing have enabled machines to learn, including deep learning networks that copy the human brain and learning algorithms that learn from prior knowledge that is being used in medical practice [7]. This paper has the following sequence: Section II represents the ML algorithm and techniques. Various classifications and hybrid approaches are discussed in Section III on the prediction of CVD. Section VI focuses on performance measure indices. Section V contains limitations of existing models, and section XI contains future work and a conclusion.

1. **ML ALGORITHM AND TECHNIQUES**

Machine learning is an emerging technology in all related fields such as Medical Science, Stock Market, Automation, Image, Robotics Processing, etc. Algorithms for machine learning build mathematical models to allow predictions or decisions to be made without having to be explicitly programmed. These models combine computer science and statistics, which are done using algorithms that learn from historical data. Machine learning uses stored data, features, and attributes to generate results [8].

* 1. **MACHINE LEARNING:** ML algorithms use mathematical concepts to find underlying patterns in data, map embedded patterns, and establish correlations. Machine learning techniques consist of a set of algorithms that can recognize patterns, classify data, split data, and predict future events based on past experience. the machine learning model works on either classification either regression or both but disease detection is a classification-based problem. Various machine learning algorithms are available for various tasks which are described below.

1. **SVM**: Classification and regression problems can both be solved using Support Vector Machines, but classification is more common. The SVM algorithm works on the principle of dividing classes and generating hyperplanes between classes according to datasets. The data points are considered by the SVM to create a decision boundary, which divides the two classes based on the data points. This is how results are generated on the basis of the data partition [9,10,11,12]
2. **KNN:** A classification principle is used by the k-Nearest Neighbor algorithm.This method assumes that the case/data is comparable to what is already accessible, and it categorizes the case into the category that is closest to what is presently available. The Nearest Neighbors means one instance which is most similar to an existing database, the algorithmidentifies those instance that is based on classification and regression [13-14].
3. **NAIVE BAYES:** In the NB classifier, the Bayesian theorem is used to create a set of algorithms. A classification algorithm is made up of a bunch of algorithms based on the notion that each pair of features to be categorized is unique. The existence of one attribute in a class has no link to the existence of any other attribute in that class, according to a Naive Bayes classifier [15, 16].
4. **RANDOM FOREST:** In order to solve classification and regression issues, Random Forests machine learning algorithms are used. This method generates a decision tree that provides better results for the same number of attributes and large data sets. Machine learning techniques such as Random Forests are used for classification and regression tasks. This method generates a decision tree that provides better results for the same number of attributes and large data sets. The Random Forest technique divides the dataset into smaller parts and uses the majority vote for classification and the average for regression. Thus the decision tree is constructed on the basis of average and majority [17-18].
5. **LR:** Linear Regression algorithm used for a regression problem. The algorithm works on two variables where one is the independent variable and the other is the dependent variable and the algorithm shows the linear relationship between these variables. The mathematical notation is Y=MX+B, where the independent variable is plotted on the X-axis and the dependent is plotted on the Y-axis. [19–20].
6. **LOGISTIC REGRESSION:** The logistic regression algorithm is similar to linear regression. This algorithm also works on the independent and dependent variables except that how they are used. The algorithm is basically used for solving classification problems. For predicting the result, the algorithm uses a mathematical sigmoid function. The threshold value of logistic regression defines the probability that it will either be zero or one. Logistic regression requires a value between 0 and 1, which cannot go beyond this limit, so it forms an S curve. It is also called the logistic function or the Sigmoid curve [21,22,23].
   1. **ENSEMBLE TECHNIQUES:** The technique of using more than one machine learning algorithm jointly is known as the Ensemble technique. Ensemble techniques are being used for more consistent and better results [24]. Bagging and boosting are two types of ensemble techniques that are described below.
7. **BAGGING:** Bagging is an ensemble machine-learning technique known as bootstrap aggregation. In the bagging ensemble technique, the dataset is divided into sub-datasets, and results are extracted from all these sub-datasets by one or many machine learning algorithms. The final prediction is generated by averaging the results obtained from all the sub-datasets [25].
8. **BOOSTING:** Developing a strong classifier by combining a series of weak classifiers is known as the boosting ensemble technique. In this technique, many models are linked to each other, the first model which produces wrong results is corrected by the second model, and this sequence goes on till the complete true result is obtained. The learning model involved in the whole process can be the same algorithm or can be different algorithms [25,26,27,28].
9. **LITERATURE REVIEW**

Machine learning is a growing field of medical sciences and healthcare. Algorithms and methodologies related to machine learning have been extensively used in the prediction of diseases. UCI Disease datasets are freely available in the UCI Machine Learning Repository for understanding and analysis. These datasets are used by researchers in the machine-learning community to test algorithms empirically [29]. For cardiovascular disease prediction, Yar Muhammad et al. used the Hungarian and Cleveland heart disease datasets. For cardiovascular disease prediction, they have used various machine learning such as NB, DT, RF, SVM, AB, GB, LR, and KNN. FCBF, LASSO, MRMR, and RELIEF were used as four feature selection algorithms to remove unusable, noisy, and irrelevant data from the dataset and they got 93.36% accuracy with the gradient boost algorithm [30]. In the order of CVD disease prediction, Kumar et al. proposed a quantum-enhanced machine learning technique. In this technique, they used the UCI Machine Learning Repository dataset with 14 features. They used four different quantum-based machine learning algorithms Quantum Decision Tree, Quantum Random Forest, Quantum K-Nearest Neighbors, and Quantum Gaussian Naive Bayes, and analyzed the prediction accuracy on the confusion matrix [31]. For effective CDV prediction, Ghosh et al. used an ensemble machine-learning technique. In this model, they used bagging techniques with various machine learning algorithms that are AdaBoost Boosting (ABB), Decision Tree Bagging (DTB), Random Forest Bagging (RFB), K-Nearest Neighbors Bagging (KNNB), and Gradient Boost Boosting (GBB). They also used two feature extraction techniques such as LASSO and Relief. They combined five different UCI datasets into a single dataset. They achieved 99% accuracy with the extracted datasets [32]. A hybrid random forest linear model (HRFLM) was proposed by Mohan et al. In this model, they used 297 patient records and 13 attributes for CVD prediction. They used different ML algorithms for disease prediction they got 88.70% accuracy on hybrid random forest [33]. Cheng et al. proposed a hybrid method for the detection of cardiovascular disease. In this technique, they used a series of ML algorithms that is KNN, DT, SVM, and Naive Bayes. They identify the hidden pattern between datasets and got 86.80% accuracy on this technique [34]. In order to improve prediction accuracy, an ensemble technique was applied to datasets. The detection of the risk of heart disease has been improved through the application of boosting and bagging methods. In this study, a hybrid model was built by using Bayes Net, NB, C-4.5, Multilayer Perceptron, RF, and PART as classifiers and this model achieved an 85.48% accuracy score [35]. For the prediction of the CVD model, Kim et al. used national health insurance service health screening datasets. In this model, they Combined two different data of CVD patients and non-CVD patients over the age of 45. They observed that the previous history of cardiovascular disease provides a significate role in the prediction of CVD. on these datasets they got high accuracy on XGB, GB, and RF algorithms [36]. An ensemble technique is proposed by Javid et al. In this model, they used the majority vote method for the prediction of CVD. They used a weak classifier and generate a strong classifier to enhance the accuracy of prediction. They increased 2.1% the accuracy of the existing model [37]. Kundu et al. described a statistical analysis of cardiovascular disease. They analyzed 65552 patient data which have aged 45 or older then and described the various risk factors and features of CVD. For this study, they accessed data from the Longitudinal Aging Study (LASI) in India which is the first data set of the first longitudinal aging study. They examine the various factors that cause CVD and see that family history and genetics are the main influencing factors for CVD. The result of this study is that cardiovascular disease took the leading cause of death in India [38]. Nadakinamani et al. Developed a Heart Disease Detection and or Prediction System based on the dataset set Statlog and Hungarian from the UCI Machine Learning dataset. They used seven different and latest machine learning algorithms (i.e., NB, LR, REP Tree, M5P Tree, Random, JRIP, and J48). A random forest algorithm was found to be most accurate when used with two different datasets of Hungarian and Statlog with 14 disease features. [39]. Hossain et al. described a new machine learning-based approach for predicting cardiovascular disease of type 2 diabetes patients. In this model, they used six different machine learning algorithms as LR, K NN, SVM, DT, RF, and NB on the data set and analyzed the result. For this model, they used Australia’s CBHS health fund patient data [40]. Yang et al. Completed the National High-Risk Screening Program Centre project to analyze heart disease. The program run from 2014 to 2016 and collected disease data from a variety of Zhejiang provinces in China. Data from 101056 patients were taken in this program, out of which 29930 were highly affected by cardiovascular disease. For the CVD prediction model, 30 different types of features were taken and seven different machine learning algorithms were applied. In the end, they got 78% accuracy on the random forest algorithm [41]. Krittanawong et al. analyzed the ML algorithm on different aspects of heart disease such as heart failure, coronary artery disease, cardiac arrhythmias, and stroke. They analyzed the accuracy of various heart problems with different ML algorithms and observed that support vector machine algorithms outperform other algorithms in these areas [42]. An investigation process applied by Weng et al. to whether machine learning can improve cardiovascular issue detection using routine clinical data is described. For this research, they analyzed Clinical Practice Research Datalink (CPRD) data records from 700 UK families. They used multiple machine learning methods to examine data from 378,256 patients. where they took 22 attributes to predict the result and got 95% accuracy on the prediction [43]. A variety of learning algorithms including Bayes Net, J48, KNN, multilayer perceptrons, Nave Bayes, random trees, and random forest were utilized by Ashraf et al. they found that the J-48 algorithm gained 70.77% accuracy [44]. A systematic review presented by Aleksei et al in the field of CVD prediction models. Researchers studied 27 different research papers on prediction models for cardiovascular disease using machine learning and found that all of the methods are based on mortality risk prediction. They believe that ensemble machine learning techniques, bagging, and boosting provide the best accuracy and performance for predicting CVD, and they suggest that data pre-processing, feature selection, dimensionality reduction, and effective attributes will provide effective results for CVD prediction [45]. Wei Chen Sun et al. In their analysis of machine-learning algorithms, they found that the SVM algorithm performs best when it comes to classification problems [46]. Researchers analyzed various machine-learning algorithms for CVD prediction. Disease prediction is a classification problem, there are two possibilities in disease prediction such as disease present or not. The result is based on the disease datasets hence it is very essential that data should be filtered and classified. Most techniques used the UCI dataset that is predefined and structured. the prediction result may be affected due to the manner of the dataset. Below table 1.0 describes the datasets used by the above ML techniques for disease prediction.

|  |  |  |
| --- | --- | --- |
| **Sl. No.** | **Type of dataset** | **Dataset** |
| 01. | UCI Machine Learning  Data Repository | Cleveland [28, 30] |
| 02. | Hungarian [30, 39] |
| 03. | Switzerland [30] |
| 04. | VA Long Beach [30] |
| 05. | Statlog [30, 39] |
| 06. | Research group dataset | Svetlana Ulianova [46] |
| 07. | Patient dataset health screening | National Health Insurance Service [36] |
| 08. | Patient of UK family’s dataset | Clinical Practice Research Datalink (CPRD) [43] |
| 09. | Centre project dataset | National High-Risk Screening Program [41] |

**Table 1.0 Datasets of CVD prediction model**

There are multiple techniques and algorithms available for classification and regression problems. The performance and/or Accuracy are evaluated by the Confusion matrix on some kind of parameters. The evaluation method of ML algorithms and techniques are described in the next section.

1. **PERFORMANCE EVALUATION**

The performance of machine learning algorithms and/or techniques is assessed on a confusion matrix. The performance of the algorithm is calculated by F-Score, Zero Error Rate, Precision, Recall, ROC Curve, Specificity, Prevalence, etc. The confusion matrix is responsible for measuring and comparing performance [47].

* 1. **CONFUSION MATRIX:** To evaluate the performance of the machine learning algorithm Makridakis et al presented a binary class confusion matrix and a Multiclass confusion matrix. The basic fundamental of confusion binary metrics is they calculate four different conditions such as TP, FP, TN, and FN (i.e., True Positive, False Positive, True Negative, & False Negative). In multi-class confusion metrics, all positions are increased in the same order [48]. Following is a table describing the binary class classification metric and the multi-class classification metric (i.e., table 2.0 table 3.0).

|  |  |  |  |
| --- | --- | --- | --- |
| **Binary Class Confusion Matrix** | | Predictive Class | |
| Positive | Positive |
| Actual  class | Positive | **TP** | **FN** |
| Negative | **FP** | **TN** |

**Table 2.0 Binary class confusion matrix**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Multiclass confusion matrix** | | Predictive Class | | | |
| **C1** | **C2** | **….** | **Cn** |
| Actual  Class | **C1** | **C1,1** | **FP** | **.…** | **C1, n** |
| **C2** | **FN** | **TP** | **….** | **FN** |
| **.…** | **….** | **…..** | **….** | **….** |
| **Cn** | **Cn,1** | **FP** | **….** | **Cn, n** |

**Table 3.0 Multiclass confusion matrix**

Where:

1. **TP = True Positive**, the patient's cardiac ailment was correctly recognized by the model.
2. **TN = True Negative**, which properly recognized the opposing type of patients, such as those without cardiac issues.
3. **FP = False Positive**, Heart disease patients who were incorrectly identified by the model, i.e., when non-heart disease patients were misidentified as heart disease patients
4. **FN = False Negative** occurs when the model incorrectly classifies a patient as having heart disease but the model does not recognize it as a heart disease [49, 50].
   1. **ACCURACY:** The accuracy of the confusion matrix is calculated by dividing the properly-recognized data instance by the total number of data instances. The equation of accuracy is as follows:

**ACCURACY = (TP+TN)/(TP+FP+FN+TN)**

* 1. **PRECISION:** The number of correctly anticipated positive observations divided by the total expected positive observations is known as precision. Precision refers to the quality of the results obtained. The equation of Precision is as follows:

**PRECISION = TP/(TP+FP)**

* 1. **RECALL:** The recall value is calculated by dividing correctly predicted positive observations by all observed observations. Sensitivity is also known as the ability to respond. The equation of Recall is as follows:

**RECALL = TP/(TP+FN)**

* 1. **F-SCORE:** F1 Score is measured using the arithmetic average of Precision & Recall. When calculating the F1 score, precision and recall are taken into account, as detailed below.

**F1 SCORE = 2\*(RECALL \* PRECISION) / (RECALL + PRECISION)**

As the specialty of confusion matrix, we can calculate various performance indices (i.e., Negative predictive value, False Negative Rate, Mathews Correlation coefficient, Prevalence Threshold, Informedness, Threat Score, Fowlkes-Mallow’s index, and so on.) [51, 52, 53, 54].

1. **LIMITATIONS OF EXISTING TECHNIQUES**

Almost everyone is concerned about their health nowadays after COVID-19. Researchers facing major challenges in developing better diagnostic and clinical facilities. Machine learning and/or Deep learning is a growing branch in the health sector and it is being seen by researchers as a better solution for various health problems. this study, review cardiovascular disease (CVD) prediction, model, and techniques. for the classification problems, the researchers used various kinds of ML algorithms and ensemble techniques. To predict CVD most researchers used the UCI Machine Learning Repository data set for CVD prediction and very few researchers used the clinical practice dataset. The UCI dataset used by researchers is already structured and filtered [28,30,39,46]. Hence the much better accuracy score provided by the ML algorithm and techniques is a huge loophole. Accuracy scores will be affected on real medical datasets and/or daily clinical datasets instead of UCI datasets. Therefore, it is very important to use more diverse real word datasets for model training. Another issue is there are no fixed guidelines and/or architecture for model training, parameter tuning, model section, and data splitting (train, test). Therefore, arbitrary model selection, parameter tuning, and data segmentation (train, test) will affect disease prediction models to a great extent, which is a major challenge for disease prediction models. The arbitrary model selection, parameter tuning, feature selection, and data splitting increase the risk of overfitting and underfitting problems [55, 56]. The accuracy is very much affected due to overfitting and the underfitting problem of the machine learning model. Most studies reported only technical aspects due to a lack of physician supervision, so the results were too skewed. In addition, the proposed technique is also overfitted because of its enormous complexity, leading to unpredictable results due to its overfitting. Hence the type of custom-built algorithm cannot be classified due to ambiguity. Thus, the issues presented in disease prediction models and/or techniques cannot be ignored [37].

1. **CONCLUSION**

In clinical practice, machine-learning applications should be made practical and acceptable. Machine learning will be a realistic option for enhancing clinical practice prediction/detection of disease risk as computing capacity increases in healthcare systems. due to strong computational power and prediction ability, ML is a powerful tool for medical science and healthcare. Disease diagnosis and/or detection is the classification problem and there are multiple techniques and algorithms available hence the architecture and proper guidelines is very essential for using ML techniques. Machine learning algorithm works on datasets and based on the previous record the algorithm predicts. Hence real-world datasets such as routine checkups, hospital records, case studies, and so on will provide better results [57]. The basic need of any machine learning algorithm is model selection, parameter tuning, feature selection, training, and testing data, the effectiveness of the result of the algorithm depends on how and which data sets are used in the algorithm. The basic need of any machine learning algorithm is model selection, parameter tuning, feature selection, training, and testing data, the effectiveness of the result of the algorithm depends on how and which data sets are used in the algorithm. In this review article, we analyzed different machine learning techniques, analyzed their results, and analyzed the datasets used. Several problems in existing models and techniques have also been clarified, which will prove to be very useful in future disease diagnosis models. We hope this review will prove to be a foundation stone under the guidance of researchers and will help researchers effectively in CVD prediction.

**Abbreviations:**

DLT: Deep Learning Techniques

DL: Deep learning

DNN: Deep neural network

ML: Machine learning

AI: Artificial intelligence

CVD: cardiovascular disease

ANN: artificial neural network

UCI: University of California, Irvine

TP: True Positive

FP: False Positive

TN: True Negative

FN: False Negative

FCBF: Fast Correlation-Based Filter

MRMR: Maximum Relevance Minimum Redundancy

LASSO: Least Absolute Shrinkage and Selection Operator

CHD: Coronary artery disease

CNN: convolutional neural network

NB: Naive Bayes

LR: Logistic Regression

RF: Random Forest

XGB: Extreme Gradient Boost

KNN:  k-nearest Neighbors

DT: Decision Tree

SVM: Support Vector Machine

**Data Availability**

Data used to support the findings of this study will be provided by the corresponding author upon request.

**Conflicts of Interest**

This publication contains no conflict of interest on the part of any of the authors since this work did not receive any financial assistance or grant, partially or fully.

**REFERENCE**

1. Zhang L, Yang H and Yang P “The Correlation between Type 2 Diabetes Mellitus and Cardiovascular Disease Risk Factors in the Elderly”, Hindawi Applied Bionics and Biomechanics (Research Article), https://[doi.org/10.1155/2022/4154426](https://doi.org/10.1155/2022/4154426), PP. 1-7, 2 January-2022.
2. Wang Y, Fai WanI ER., Mak LL, Kay Ho M, Yee Chin W, Tak Yu EY and lo Kuen Lam C, “The association between trajectories of risk factors and risk of cardiovascular disease or mortality among patients with diabetes or hypertension: A systematic review”, Plos One (Research Article), https://[doi.org/10.1371/journal.pone.0262885](https://doi.org/10.1371/journal.pone.0262885), PP. 1-15, January-2022.
3. “Cardiovascular diseases (CVDs)”, Access date 18 February-2022<https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases>
4. Roth GA. et al. “Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019”, Journal of the American college of cardiology, vol. 76, Issue- 25, PP. 1-40, December-2020.
5. Kumar AS and Sinha A, “cardiovascular disease in India: A 360-degree overview”, ScienceDirect, DOI: [10.1016/j.mjafi.2019.12.005](https://dx.doi.org/10.1016%2Fj.mjafi.2019.12.005), Vol. 76, Issue-1, PP. 1-3, January 2020.
6. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, and Callaway CW, “2021 heart disease and Stroke Statistics Update Fact Sheet At-a-Glance” American Heart Association, Vol. 143, Issue-8, PP. 1-6, January-2021.
7. Mathur P, Srivastava S, Xu X and Mehta JL, “Artificial Intelligence, Machine Learning, and cardiovascular disease”, Clinical Medicine Insights: Cardiology, DOI: [10.1177/1179546820927404](https://dx.doi.org/10.1177%2F1179546820927404), Vol. 9, Issue-1, PP. 1-9, September-2020.
8. Jordan MI and Mitchell TM, “Machine learning: Trends, perspectives, and prospects”, American Association for the Advancement of Science, [DOI: 10.1126/science.aaa8415](https://doi.org/10.1126/science.aaa8415), Vol. 349, Issue-6245, PP. 1-7, Jul-2015.
9. Srivastava DK and Bhambhu L, “Data classification using support vector machine”, Journal of Theoretical and Applied Information Technology, Vol. 12, Issue-1, PP. 1-7 February-2010.
10. Bhavsar H and Panchal MH, “A Review on Support Vector Machine for Data Classification”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol. 1, Issue-10, PP. 1-5, December-2012.
11. Suresh T, Assegie TA , Rajkumar S, and Kumar NK, “A hybrid approach to medical decision-making: diagnosis of heart disease with machine-learning model”, DOI: 10.11591/ijece.v12i2.pp1831-1838, International Journal of Electrical and Computer Engineering (IJECE), Vol. 12, Issue 2, PP. 1-8, April-2022
12. Aswini J, Yamini B, Jatothu, Nayaki KS, and Nalini M ,“An efficient cloud-based healthcare services paradigm for chronic kidney disease prediction application using boosted support vector machine”, <https://doi.org/10.1002/cpe.6722>, Currency and Computation, Vol. 34, Issue 10, PP. 1-8, December 2021.
13. Cunningham P and Delany SJ, “k-Nearest Neighbour Classifiers - A Tutorial” [ACM Computing Surveys](https://dl.acm.org/toc/csur/2022/54/6), https://[doi.org/10.1145/3459665](https://doi.org/10.1145/3459665), Vol. 54, Issue-6, PP. 1-25, July-2022.
14. Bansal M, Goyal A and, Choudhary A, “A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning”, <https://doi.org/10.1016/j.dajour.2022.100071>, [Decision Analytics Journal](https://www.sciencedirect.com/journal/decision-analytics-journal), Vol. 3, Issue-, PP. 1-21, June-2022.
15. Rahat A M, Kahir A, and Masum AKM, “Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset”, DOI**:**[10.1109/SMART46866.2019.9117512](https://doi.org/10.1109/SMART46866.2019.9117512), 8th International Conference on System Modeling & Advancement in Research Trends (IEEE Access), PP. 1-5, November-2019.
16. Yilmaz AB, Taspinar YS, and Koklu M, “Classification of Malicious Android Applications Using Naive Bayes and Support Vector Machine Algorithms” International Journal of intelligent systems and applications in engineering, Vol. 10, Issue-2, PP. 1-6, May-2022.
17. Kurdi FT, Amakhchan W, and Gharineiat Z, “Random Forest Machine Learning Technique for Automatic Vegetation Detection and Modelling in LiDAR Data”, DOI: [10.19080/IJESNR.2021.28.556234](http://dx.doi.org/10.19080/IJESNR.2021.28.556234), International Journal of Environmental Sciences & Natural Resources, Vol. 28, Issue-2, PP. 1-3, June-2021.
18. Rostami M, and Oussalah M, “A novel explainable COVID-19 diagnosis method by integration of feature selection with random forest”, <https://doi.org/10.1016/j.imu.2022.100941>, Informatics in Medicine Unlocked, Vol. 30, Issue-, PP. 1-15, April-2022.
19. Maulud DH and Abdulazeez AM, “A Review on Linear Regression Comprehensive in Machine Learning”, Journal of Applied Science and Technology Trends, Vol. 1, Issue-4, PP. 1-8, December- 2020.
20. Zhang D, Khalili A, and Asgharian M “Post-model-selection inference in linear regression models: An integrated review”, <https://doi.org/10.1214/22-SS135>, Statist. Surv. 16 86 - 136, 2022, Vol. 16, Issue-, PP. 1-51, April 2021.
21. Lee H and Kim H, “Logistic Regression and Least Absolute Shrinkage and Selection Operator”, Cardiovascular Prevention and Pharmacotherapy, https://[doi.org/10.36011/cpp.2020.2.e15](https://doi.org/10.36011/cpp.2020.2.e15), [Vol. 2](https://www.e-jcpp.org/articles/current.php?vol=2&no=4), Issue- 4, PP. 1-5, March-2020.
22. Sarker IH “Machine Learning: Algorithms, Real‑World Applications, and Research Directions”, https://doi.org/10.1007/s42979-021-00592-x, PP. 1-21, March-2021.
23. Ramalingam VV, Dandapath A, and Raja MK, “Heart disease prediction using machine learning techniques: a survey”, DOI:[10.14419/ijet.v7i2.8.10557](http://dx.doi.org/10.14419/ijet.v7i2.8.10557), International Journal of Engineering & Technology, Vol. 7, Issue-2, PP. 1-4, March-2018.
24. Singh R and Pal S, “Machine Learning Algorithms and Ensemble Technique to Improve Prediction of Students Performance”, International Journal of Advanced Trends in Computer Science and Engineering, DOI:[10.30534/ijatcse/2020/221932020](http://dx.doi.org/10.30534/ijatcse/2020/221932020), Vol. 9, Issue-3, PP. 1-7, Jul-2020
25. Qutub A, Al-Mehmadi A, Al-Hssan M, Aljohani R and Alghamdi HS, “Prediction of Employee Attrition Using Machine Learning and Ensemble Methods”, International Journal of Machine Learning and Computing, DOI:[10.18178/ijmlc.2021.11.2.1022](http://dx.doi.org/10.18178/ijmlc.2021.11.2.1022), Vol. 11, Issue- 2, PP. 1-5, March-2021.
26. Lee H and Kim H, “Logistic Regression and Least Absolute Shrinkage and Selection Operator”, Cardiovascular Prevention and Pharmacotherapy, DOI:[10.36011/cpp.2020.2.e15](http://dx.doi.org/10.36011/cpp.2020.2.e15), [Vol. 2](https://www.e-jcpp.org/articles/current.php?vol=2&no=4), Issue- 4, PP. 1-5, March-2020.
27. Buhlmann P, “Bagging, Boosting, and Ensemble Methods”, Handbook of Computational Statistics, DOI: 10.1007/978-3-642-21551-3\_33, PP. 985-1022, December-2011.
28. Jafarzadeh H, Mahdianpari M, Gill E, Mohammadimanesh F, and Homayouni S, “Bagging and Boosting Ensemble Classifiers for Classification of Multispectral, Hyperspectral, and PolSAR Data: A Comparative Evaluation”, https://[doi.org/10.3390/rs13214405](https://doi.org/10.3390/rs13214405), Remote Sens (MDPI), Vol. 13 Issue- 21, PP. 1-22, November-2021.
29. “UCI Machine Learning Repository”, Access date 18 February-2022 <https://archive.ics.uci.edu/ml/about.html>.
30. Muhammad Y, Tahir M, Hayat M, and Chong K, “Early and accurate detection and diagnosis of heart disease using intelligent computational model”, https://doi.org/10.1038/s41598-020-76635-9, Published by Scientific Reports, [10](https://www.e-jcpp.org/articles/current.php?vol=2&no=4), Issue- 4, PP. 1-18, November 2020.
31. Kumar Y, Koul A, Sisodia PS, Shafi J, Verm V, Gheisari M, and Davoodi MB, “Heart Failure Detection Using Quantum-Enhanced Machine Learning and Traditional Machine Learning Techniques for Internet of Artificially Intelligent Medical Things”, https://[doi.org/10.1155/2021/1616725](https://doi.org/10.1155/2021/1616725), Wireless Communications and Mobile Computing (Research Article), Vol. 2021, Issue-, PP. 1-16, December-2021.
32. Ghosh P, Azam S, Jonkman M, Karim S, Shamrat FMJM, Ignatious E, Shultana S, Beeravolu AR, and De Boer AF, “Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques”, DOI:[10.1109/ACCESS.2021.3053759](https://doi.org/10.1109/ACCESS.2021.3053759), IEEE ACCESS, Vol. 09, Issue-, PP. 1-23, February-2021.
33. Mohan SK, Thirumalai CH, and Srivastava G, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques”, DOI:[10.1109/ACCESS.2019.2923707](https://doi.org/10.1109/ACCESS.2019.2923707), IEEE ACCESS, Vol. 07, Issue-, PP. 1-13, June-2019.
34. Cheng CA and Chiu HW, “An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database”, DOI:[10.1109/EMBC.2017.8037381](https://doi.org/10.1109/EMBC.2017.8037381), 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), PP. 2566–2569, July-2017.
35. Latha CBC and Jeeva SC, “Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques”, https://[doi.org/10.1016/j.imu.2019.100203](https://doi.org/10.1016/j.imu.2019.100203), Informatics in Medicine Unlocked, vol. 16, Issue-, PP. 1-9, January-2019.
36. Kim JO, Jeong YS, Kim JH, Lee JW, Park D, and Kim HS, “Machine Learning-Based Cardiovascular Disease Prediction Model: A Cohort Study on the Korean National Health Insurance Service Health Screening Database”, https://[doi.org/10.3390/diagnostics11060943](https://doi.org/10.3390/diagnostics11060943), [Machine Learning and Artificial Intelligence in Diagnostics](https://www.mdpi.com/journal/diagnostics/sections/artificial_Intelligence) (MDPI), Vol. 11, Issue- 6, PP-1-12, May-2021.
37. Javid I, Khalaf A and Ghazali R, “Enhanced accuracy of heart disease prediction using machine learning and recurrent neural networks ensemble majority voting method”, DOI:[10.14569/IJACSA.2020.0110369](http://dx.doi.org/10.14569/IJACSA.2020.0110369), International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 11, Issue- 3, PP. 1-12, March-2020.
38. Kundu J and Kundu S, “cardiovascular disease (CVD) and its associated risk factors among older adults in India: Evidence from LASI Wave 1”, https://[doi.org/10.1016/j.cegh.2021.100937](https://doi.org/10.1016/j.cegh.2021.100937), Clinical Epidemiology and Global Health (Elsevier), Vol. 13, Issue-, PP.1-5, January- 2021.
39. Nadakinamani RG, Reyana A, Kautish S, Vibith AS, Gupta Y, Abdelwahab SF, and Mohamed AW, “Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques”, https://[doi.org/10.1155/2022/2973324](https://doi.org/10.1155/2022/2973324), Computational Intelligence and Neuroscience (Research Article), Vol. 2022, Issue-, PP. 1-13, February-2022.
40. Hossain ME, Uddin S and Khan A, “Network analytics and machine learning for predictive risk modeling of cardiovascular disease in patients with type 2 diabetes”, https://[doi.org/10.1016/j.eswa.2020.113918](https://doi.org/10.1016/j.eswa.2020.113918), Expert Systems With Applications, [Vol. 64, Issue-](https://www.sciencedirect.com/journal/medical-journal-armed-forces-india/vol/76/issue/1), PP. 1-13 September-2020.
41. Yang L, Wu H, Jin X, Zheng P, Hu S, Xu X, Yu W, and Yan J, “Study of cardiovascular disease prediction model based on random forest in eastern China”, https://doi.org/10.1038/s41598-020-62133-5, Scientific Reports, [Vol. 10, Issue-](https://www.sciencedirect.com/journal/medical-journal-armed-forces-india/vol/76/issue/1), PP. 1-8, March-2020.
42. Krittanawong C, Virk HUH, Bangalore S, Wang Z, Johnson KW, Pinotti R, Zhang H, Kaplin S, Narasimhan B, Kitai T, Baber U, Halperin JL, and Tang WHW, “Machine learning prediction in cardiovascular diseases: a meta-analysis”, https://doi.org/10.1038/s41598-020-72685-1, Scientific Reports, [Vol. 10, Issue-](https://www.sciencedirect.com/journal/medical-journal-armed-forces-india/vol/76/issue/1), PP. 1-11, September-2020.
43. F.Weng S, Reps J, Kai J, Garibaldi JM, and Qureshi N, “Can machine-learning improve cardiovascular risk prediction using routine clinical data?”, https://[doi.org/10.1371/journal.pone.0174944](https://doi.org/10.1371/journal.pone.0174944), PLOS ONE, [Vol. 12, Issue-](https://www.sciencedirect.com/journal/medical-journal-armed-forces-india/vol/76/issue/1)4, pp-1-15, April-2017.
44. Ashraf M, Ahmad SM, Ganai NA, Shah RA, Zaman M, Khan SA, and Shah AA, “Prediction of Cardiovascular Disease Through Cutting-Edge Deep Learning Technologies: An Empirical Study Based on TENSORFLOW, PYTORCH and KERAS”, DOI: 10.1007/978-981-15-5113-0\_18, [International Conference on Innovative Computing and Communications](https://link.springer.com/book/10.1007/978-981-15-5113-0), PP. 239-255 January 2021
45. Dudchenko A, Ganzinger M and Kopanitsa G, “Machine Learning Algorithms in Cardiology Domain: A Systematic Review”, DOI**:**[10.2174/1875036202013010025](http://dx.doi.org/10.2174/1875036202013010025), The Open Bioinformatics Journal, vol. 13, Issue-, PP. 1-16, February-2020.
46. Weicheng S, Ping Z, Zilin W, and Dongxu L, “Prediction of Cardiovascular Diseases based on Machine Learning”, <https://doi.org/10.52810/TIOT.2021.100035>, ASP Transactions on internet of things, vol. 1, Issue- 1, PP. 1-6, April-2021.
47. Orozco-Arias S, Pina JS, Tabares-Soto R, Castillo-Ossa LF, Guyot R, and Isaza G, “Measuring Performance Metrics of Machine Learning Algorithms for Detecting and Classifying Transposable Elements”, <https://doi.org/10.3390/pr8060638>, Processes Research Article (MDPI), Vol.8, Issue-6, PP. 1-20 May 2020.
48. Markoulidakis I, Rallis I, Georgoulas I, Kopsiaftis G, Doulamis A, and Doulamis N, “Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem”, <https://doi.org/10.3390/technologies9040081>, Technologies (MDPI) Article, PP. 1-13, November 2021.
49. Vakili M, Ghamsari M and Rezae M, “Performance Analysis and Comparison of Machine and Deep Learning Algorithms for IoT Data Classification”, <https://doi.org/10.48550/arXiv.2001.09636>, arXivLabs Cornell University [Vol. 1, Issue-](https://www.sciencedirect.com/journal/medical-journal-armed-forces-india/vol/76/issue/1), PP. 1-13, January- 2020.
50. Sokolova M, Japkowicz N, and Szpakowicz S, “Beyond Accuracy, F-score and ROC: a Family of Discriminant Measures for Performance Evaluation”, DOI: 10.1007/11941439\_114, Australasian Joint Conference on Artificial Intelligence, PP 1-7, January-2006.
51. Goutte C and Gaussier E, “A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation”, DOI: 10.1007/978-3-540-31865-1\_25, Proceedings of the 27th European Conference on Advances in Information Retrieval Research, PP 1-16, April-2005.
52. [Powers](https://arxiv.org/search/cs?searchtype=author&query=Powers%2C+D+M+W) DMW, “Evaluation: from precision, recall, and f-measure to roc, informedness, markedness & correlation”, <https://doi.org/10.48550/arXiv.2010.16061>, Journal of Machine Learning Technologies Vol. 2, Issue-1, PP. 1-27, October-2020.
53. Yacouby R and Axman D, “Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models”, DOI: [10.18653/v1/2020.eval4nlp-1.9](http://dx.doi.org/10.18653/v1/2020.eval4nlp-1.9), Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems, PP 1-13, January-2020.
54. Gong M, “A novel performance measure for machine learning classification”, International Journal of Managing Information Technology (IJMIT), Vol.13, Issue-1, PP. 1-19 February-2021.
55. engio Y. Practical Recommendations for Gradient-Based Training of Deep Architectures. In: Montavon G, Orr GB, Mu¨ller K-R, eds. Neural Networks: Tricks of the Trade: Second Edition. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012: 437–78
56. Liu B, Wang S, Dong Q, Li S, Liu X. Identification of DNA-binding proteins by combining auto-cross covariance transformation and ensemble learning. IEEE Trans Nanobioscience 2016; 15(4): 328–44
57. Prajapati AK and Singh UK “An empirical study of machine learning (ml) algorithms in the perspective of cardiovascular disease (cvd) prediction”, SSRN Preprints, <http://ssrn.com/abstract=4117242>, May 2022.