Review Article

International Journal of Advanced Computer Research, Vol 12(58)
ISSN (Print): 2249-7277   ISSN (Online): 2277-7970
http://dx.doi.org/10.19101/IJACR.2021.1152057

A review and methodological analysis of cardiovascular disease prediction and detection

Bhavana Jauhari*, Animesh Dubey and Mohd Zuber
Madhyanchal, Professional, University (MPU), Bhopal, India

Received: 30-November-2021; Revised: 11-January-2022; Accepted: 15-January-2022

Abstract
According to the world health organization (WHO) cardiovascular diseases (CVDs) are the leading cause of death globally. This paper examines and explores different methodological contribution in the detection and prediction of CVDs. This paper mainly covers the ways of applicability of the approaches, domain of applicability, results, advantages, challenges and limitations. It also investigated the datasets from different repository. The exploration clearly discusses the major challenges along with the suggestive measures. The results indicate the combination of approaches and it may vary according to the symptoms and its nature.

Keywords
CVD, SVM, DT, LR, KNN.

1.Introduction
Cardiovascular diseases (CVDs) are the most common cause of death around the world. CVDs claimed the lives of 17.9 million individuals worldwide in 2019, accounting for 32% of all deaths. Heart attacks and strokes are sudden occurrences caused by a blockage in the blood supply to the heart or brain. Heart attacks and strokes were responsible for 85% of these deaths [1]. Heart Disease is one of the most difficult and life-threatening human disorders. In this disease, the heart is frequently unable to pump enough blood to other regions of the body to meet the body's regular functions, and as a result, heart failure develops [2]. Early detection of a heart attack can dramatically reduce the risk of the attack occurring. Medical practitioners' daily practice generated a wealth of datasets that can be evaluated to establish the key qualities to look for when diagnosing a heart attack [3]. Many people die as a result of symptoms that were previously unidentified or simply neglected. It's past time to forecast heart disease before it strikes. Heart disease is caused by a number of factors like blood pressure, high cholesterol, alcoholic beverages, sugar, smoking, CVD, hypertensive heart and a lack of physical activity are only a few of them [1].

Machine learning is an analytical tool that is utilized when a task is vast and complex to programme. It is becoming an important tool in the health-care business to aid with patient diagnosis [4]. Surgical treatment of heart disease is difficult, especially in underdeveloped countries that lack skilled medical personnel, diagnostic equipment, and other resources needed for proper diagnosis and care of heart patients. The precise assessment of the risk of cardiac failure would aid in the prevention of serious heart attacks and enhance patient safety. When educated on appropriate data, machine learning algorithms can be useful in identifying diseases. For the purpose of comparing prediction models, public datasets on heart disease are accessible. Several research have been carried out to investigate the application of machine learning for diagnostic purposes [5,6]. Most of the studies have considered the Cleveland dataset from the University of California, Irvine (UCI) repository for the training and testing of the various machine learning models, which has been used by a number of researchers. For the purpose of comparing prediction models, public datasets on heart disease are accessible. The development of machine learning and artificial intelligence aids researchers in developing the best prediction model possible utilizing the enormous databases at their disposal [6–8]. With the help of sensors and IoTs, a new idea of chronic cardiac monitoring and remote distribution of

*Author for correspondence

smart hospital selection via mHealth is offered. Through the decision-making approach, the framework can also serve all health emergency levels of CVD patients i.e., Peril, critical, and ailing and successfully address healthcare service availability for all CVD emergency levels of patients [10]. So, in all it is found that the early detection or prediction of heart disease can be performed by with the help of various machine learning techniques, applicability of IoT sensors, deep learning, Neural Networks etc. [11–18]. Distinct studies have looked into different types of classifiers for predicting CVDs, including single and meta classifiers. When a single classifier fails to perform as expected, an ensemble classifier may be used to provide a significant improvement over the individual classifier. It was found that ensemble classifiers are more resilient and sufficient for illness prediction than individual classifiers since they train several classifiers to anticipate the final prediction outcome [19–24]. As per the research, the ensemble classifiers outperformed the individual/single classifiers in a variety of other applications fields. A variety of base classifiers and their combinations were used for the evaluation of performance of different classifiers [6, 14, 15].

As a result, we proposed the gathering of relevant data on all aspects of our field of study, training and testing the data using the discussion on machine learning method, and predicting the likelihood of a patient contracting a cardiac ailment. This study is helpful to provide a review on the various aspects associated with the heart disease and discussing the different methods and techniques suitable for the identification of heart disease. They're employed to recognize arrangements that help with forecasting and controlling mechanisms for the analysis and the part of treatment. Machine learning techniques, as outlined in multiple recent research, have the potential to provide high accuracy in categorization when compared to traditional data classification processes [16−18]. The remaining paper has been organized in the following way. Section 2 shows the literature review. Methodological analysis is shown in section 3 along with the discussion. Finally, conclusion of the review of literature has been provided.

2.Literature review

In this section literature review has been discussed and analyzed. It covers the methodological discussion and the results of the approach along with the advantages and limitations.

In 2020, Tougui et al. [25] used the UCI dataset for experimentation. There are a total of 13 features and 303 instances. Out of these 303 instances 164 are the healthy subjects and Cardiovascular illness affects 139 people. They chose to examine six popular data mining tools: Orange, Weka, RapidMiner, Knime, Matlab, and Scikit-Learn by identifying heart disease. logistic regression (LR), support vector machine (SVM), k-nearest neighbors (KNN), artificial neural network (ANN), naive Bayes (NB), and random forest (RF) were used in this study. The performance metrics selected for this comparison are confusion matrix, sensitivity, accuracy and sensitivity. ANN was found to be prominent in all. This attained the accuracy of 85.86%, sensitivity of 83.94% and specificity of 87.50%. The specificity of SVM for RapidMiner’s achieved as 94.38%.

In 2019, Kannan and Vasanthi [26] considered the dataset of Cleveland. There were 303 records in total, with 14 variables of the Cleveland. The data was divided into 70% for training and 30% of testing. RF, LR, stochastic gradient boosting (SGB), and SVM were used for the experimentation. The receiver operating characteristics (ROC) was used to evaluate the performance, with the fundamental evaluation measurements of specificity and sensitivity. They performed the comparison between area under curve (AUC) and accuracy of different used machine models. They found the AUC of 91.61% and 86.51% in case of LR, AUC of 89.53% and 80.89% in random forest, AUC of 90.70% and 84.26% in stochastic gradient boosting, and AUC of 88.26% and 79.77% in SVM. They proposed that there is a need to predict the heart disease with the help of tensor flow and deep learning algorithms with the help of more datasets.

In 2020, Tama et al. [27] considered the four different datasets. The names of datasets were Z-Alizadeh Sani, Statlog, Cleveland and Hungarian. Each dataset comprised 54, 13, 13, and 13 as the features, as well as 303, 261, 303, and 294 instances. The datasets considered the patients who had the coronary heart disease. They proposed the two-tier ensemble which further built-up on three distinct classifiers i.e., RF, gradient boosting machine (GBM), extreme gradient boosting machine (XGBoost). The performance of proposed method two-tier ensemble was compared with decision tree (DT), classification and regression tree (CART), RF, GBM, XGBoost. The performed this comparison with AUC and Friedman rank was calculated. The proposed method attained the AUC of 99.70% for the

Z-Alizadeh Sani, AUC of 93.42% for the Statlog, AUC of 85.86% for the Cleveland, AUC of 92.98% for the Hungarian. The Friedman rank achieved for the proposed method was 1.50. It's worth noting that the lower the classifier's rank, the better the classifier. The proposed model was able to outperform state-of-the-art coronary heart disease (CHD) detection based on the experimental data.

In 2022, Saboor et al. [28] performed study on the publically available datasets. They used the dataset from Cleveland Clinic foundation, StatLog and Z-Alizadeh. The considered datasets have 303 records, 270 records and again the 303 records corresponding to each selected dataset. RF, linear discriminant analysis (LDA), CART, LR, XGBoost, SVM, etc. were applied. The k-fold classification was applied to each selected model. The performance measures selected for these various models were like precision, recall, and accuracy and F measure. It was found that SVM attained the maximum accuracy of 96.72% whereas CART achieved the lowest accuracy of 83.66%. It was found that there was a problem that the classifier prediction accuracy improves as the dataset size grows larger, but after a certain point, growing dataset size has a negative impact. This has a negative impact on the predictability of the classifier.

In 2019, Gonsalves et al. [29] used the South African heart disease dataset for their study, it is a subset of a larger dataset. It has a total of 462 instances and ten attributes, 9 of which are independent factors and 1 is the variable. The dataset represents a retrospective sample of males in the Western Cape, South Africa, in a high-risk zone for heart disease-knowledge extraction based on evolutionary learning (KEEL). They applied the three different machine learning algorithms named as NB, SVM and DT on the dataset. They have considered ten-fold validation. Among all the other models the NB model turn out to be best and achieved the sensitivity of 63%, specificity of 76.16%.

In 2022, Phasinam et al. [30] employed the dataset taken from UCI repository and applied various machine learning models to predict the heart disease. They used the Iterative Dichotomiser 3 (ID3), SVM, C4.5, and NB to analyses the Cleveland dataset, which had 303 entries. The accuracy of all the models were calculated and compared. It was found that overall SVM had the better results with less errors. The accuracy achieved for SVM, C4.5, ID3 and Naïve Bayes was 92%, 85%, 78%, 70% whereas the error rate achieved was 8%, 14%, 22%, and 28%.

In 2021, Bharti et al. [31] considered the four publicly available dataset named as Cleveland, Switzerland, Long Beach V and Hungary. They contained 76 attributes. In the dataset, heart illness occurred 54.46% of the time, while no heart disease occurred 45.54% of the time. The dataset has 14 primary attributes that were used in the study. They used the LR, KNN, DT, RF, SVM and the XGBoost. Deep learning with two different ways such as sequential model and functional deep learning to predict the performance on chosen dataset. The accuracy, precision, confusion matrix, F1 score, sensitivity was used as the performance metrics for these selected algorithms. It was found that deep learning model achieved the highest accuracy of 94.2%, specificity of 83.1% and sensitivity of 82.3% whereas RF attained the lowest accuracy of 80.3%, specificity of 78.7% and sensitivity of 78.2%. It was found that if a large dataset is available, the results will be more accurate and can greatly improve deep learning and machine learning models.

In 2019, Mohan et al. [32] used the Cleveland dataset. This dataset contains the records of 303 patients. Out of which 6 records had some missing values so they were excluded and finally the data of 297 patients were considered. They proposed the hybrid model of random forest with the linear model (HRFLM). The various techniques like NB, LR, SVM, DT, deep learning, RF, GBT, VOTE and generalized linear model were used for the comparison with the proposed hybrid model. The performance measures considered were accuracy, sensitivity and specificity were used. The proposed HRFLM achieved the accuracy of 88.4%, precision of 90.1%, F-measure of 90%, sensitivity of 92.8% and specificity of 82.6%.

In 2019, Louridi et al. [33] used the Cleveland dataset for the prediction of heart disease. This dataset consisted of 13 attributes and 303 instances. They applied the SVM, KNN and NB. The performance measure considered was specificity, sensitivity, f1-score, precision, recall and accuracy as the performance measures. It was found that SVM achieved the highest accuracy of 86.8%.

In 2019, Alotaibi [34] presented an implementation of machine learning techniques on the Cleveland dataset. This dataset had 14 attributes. It was collected from Kaggle. They applied various DT, LR, RF, NB and SVM. Ten-fold cross validation was used for the validation. They used RapidMiner, Weka and Matlab. The highest accuracy of 93.19% was attained by DT whereas NB attained the lowest accuracy of 87.27%. They suggested to use the dataset of large number of patients.

In 2020, Rajdhan et al. [35] considered the Cleveland dataset available in Kaggle. It contains total 76 attributes but they used only 14 attributes. NB, DT, RF and LR were applied. RF achieved the highest accuracy of 90.16% whereas DT attained the lowest accuracy of 81.97%. The LR and NB both achieved the accuracy of 85.25%. They suggested that it will be more efficient to use it on web application and larger dataset would be more helpful.

In 2021, Kavitha et al. [36] proposed the hybrid model using machine learning approaches on the Cleveland dataset for the prediction of heart disease. The Cleveland dataset comprises of 303 instances and 14 attributes. The 70% of dataset was used as the training dataset and 30% was used as the testing dataset. The hybrid model used was the combination of DT and RF. They used the Python 3.7.6 with sklearn libraries, matplotlib, pandas etc. The accuracy of hybrid model (decision tree and random model) was compared with accuracy of individual random forest and decision tree learning techniques. They found that decision tree attained the accuracy of 79%, random forest attained the accuracy of 81% whereas the hybrid of both these techniques achieved the accuracy of 88%. They proposed that deep learning model can give better results/outcomes.

In 2020, Sharma et al. [37] considered the primary dataset collected from video of fingertip of index finger. The timespan of video was 5-35 seconds. The dataset used for the study was collected for the age group between 20 years to 50 years. These videos were recorded in Xiaomi Redmi Note 5 Pro. For the removal of noise from the dataset they used the Savitzky-Golay filter. After the collection of datasets, they proposed the modified artificial plant optimization (MAPO) for the calculation of heart rate. The preprocessing of dataset was performed and then further they applied the NB, ANN, LR and XGBoost. The performance measures considered were accuracy, precision and F1-score with all features (13) and 7 as the optimal features. It was found that the MAPO outperformed with 81.25% in case of dimensionality reduction.

In 2019, Mezzatesta et al. [38] considered the three datasets. The first dataset (D1) consisted of 522 samples. There was a total of 29 characteristics, including 8 binary attributes, 3 categorical variables, 18 discrete and continuous variables, and 3 category variables. The second dataset consists of 23 features which has 13 discrete and continuous variables, 6 categorical variables and 4 binary attributes. Lastly, the third dataset comprised of 1216 samples and had 21 features with 12 discrete and continuous variables, 2 categorical variables and 7 binary variables. They tested both linear and non-linear types of algorithms finally applied support vector machine. They found best performance on in using the non-linear support vector classifier with radial basis function kernel. They accuracy of various machine learning algorithms like LR, KNN, CART, NB, Linear support vector classifier (SVCL), support vector classifier with radial basis function kernel (SVCR), SVC with Polynomial kernel (SVCP) were applied on datasets type like CardioVascular Disease, Cardio Heart Failure, Ischemia Heart Disease, Arithmia, Cardio-CerebroVascular Disease. The performance measures like precision, accuracy, specificity, AUROC, recall F1-score and sensitivity were used with different algorithms and datasets. To optimize the performance of SVCR algorithm the grid search values provide the best results of 95.25% on Italian datasets and 92.15% in the American dataset. It was found that there is the presence of biasness.

In 2020, Ali et al. [39] employed two datasets named as Hungarian and Cleveland. These datasets were collected from UCI and other repositories. The Cleveland dataset consisted of 303 instances and 76 features. They considered the 14 and 16 features out of 76 features. On the other side, the Hungarian dataset consisted of 294 cases and having 14 features. Both these datasets were combined and comprised of 597 cases having 14 features. They evaluate the deep learning model and compared it with SVM, multi-layer perceptron (MLP), DT, RF, NB and LR. The root mean square error, mean absolute error, accuracy, F1-measure, recall, precision and recall were used. The proposed system obtained the accuracy of 98.5%.

3.Methodological analysis and discussion

Table 1 shows the methodological analysis along with the dataset. It covers the approach, results, advantages and limitations. *Figure 1* shows the methodological distribution in the detection and prediction of the heart disease.

Table 1 Methodological analysis of approaches used in heart disease classification

Sr. No.	Source	Dataset	Methods	Results	Advantage	Limitations
1	[40]	Cleveland heart disease dataset. It contains 14 numerical features and 303 instances.	NB, SVM, KNN, neural network (NN), J4.8, RF and genetic algorithm	The accuracy obtained was of 89.2%.	Different classification techniques and data mining techniques are used for extracting the optimal features.	It is recommended to consider the introducing an effective Distant heart disease prediction system to monitor and predict heart disease using patient data collected from remote devices.
2	[41]	UCI repository with 270 instances and 13 attributes.	ANN, Hybrid DT and C4.5	The hybridization of DT obtained the highest accuracy of 78.14% whereas ANN obtained the accuracy of 77.4% and C4.5 attained the accuracy of 76.66%	The number of tests required for the identification of heart disease is reduced using data mining techniques.	Using DT approaches, a web-based application or tool can be created for those living in rural areas to readily detect the condition.
3	[42]	Cleveland Clinic Foundation dataset was used which comprised of 14 attributes.	They employed fast correlation-based feature selection (FCBF). SVM, KNN, RF, RF, MLP and NB were used. Finally, ANN optimized by particle swarm optimization (PSO) which was combined with ant colony optimization (ACO) were applied.	Proposed FCBF, PSO and ACO achieved the classification accuracy of 99.65%	It is helpful in the prediction of heart disease.	The time limit for the study was less.
4	[43]	UCI repository and the data collected with the help of mobile app	Sequential minimal optimization, Bayes Net, MLP and NB	The Accuracy achieved was Sequential Minimal Optimization: 84.07%, Bayes Net: 81.11%, MLP: 77.4%, NB: 89.77%	A better understanding of the research process as a whole would help to ensure that application-based heart disease prevention research is not left behind by technological advancements.	To deliver illness apps that improve healthcare outcomes, app developers should collaborate with health care providers and researchers.
5	[44]	UCI repository. It contains 270 instances, 13 attributes. Among 270 instances,	LR, SVM, MLP, J48. 10-fold cross validation was applied.	Performance measures considered were accuracy, F1-measure, precision, recall and kappa. The accuracy obtained for these methods were: Logistic regression: 83.70%, SVM: 84.07%,	To monitor, forecast, and diagnose cardiac disease, a Cloud and IoT-based disease diagnosis	Small dataset

Sr. No.	Source	Dataset	Methods	Results	Advantage	Limitations
		150 instances had heart disease and 120 instances were having absence of heart disease.		J48: 91.48% (highest) and MLP: 78.14% (lowest)	model has been developed.	
6	[45]	National Health and Nutrition Examination Survey (NHANES) dataset	LR, SVM, RF and gradient boosting	Different cases of AU-ROC have been compared.	It aids in the automatic detection of persons at risk of diabetes and cardiovascular disease.	In order to construct more accurate models, it is necessary to investigate the usefulness of variables in electronic health records.
7	[46]	Review paper- They selected 55 studies.	Role of machine learning algorithms	Different cases of AUC have been compared.	Machine learning algorithms could be connected with electronic health record systems and used in clinical practise, especially in places with limited resources.	Firstly, it causes a lot of statistical heterogeneity. Secondly because there are no established rules for data usage, the data partition is similarly arbitrary. Third, the procedures and techniques for selecting features are arbitrary and varied. Fourth, custom-built algorithms are ambiguous. Fifth, the confusion matrix for their meta-analytic approach was not reported.
8	[47]	Cleveland database of UCI repository. There are 303 instances and 76 attributes in the collection. Only 14 of the 76 qualities are taken into account for	NB, KNN, DT and RF	The accuracy achieved by NB: 88.15%, KNN: 90.78%, DT: 80.26% and RF: 86.84%	To forecast if a patient will acquire heart disease or not.	To improve the accuracy of early cardiac disease prediction, more complicated and combined models must be implemented.

Sr. No.	Source	Dataset	Methods	Results	Advantage	Limitations
		testing.				
9	[48]	UCI machine learning repository-Cleveland Database which comprises of 303 records.	Modified deep convolutional neural network (MDCNN).	MDCNN has a 98.2% accuracy rate.	Wearable technologies can be beneficial in the healthcare profession, especially in the case of chronic heart disease.	Other feature selection algorithms and optimization Technique are required to improve the performances.
10	[49]	UCI repository	SVM, KNN, DT and linear regression	Accuracy achieved by each model. SVM: 83% DT: 79% Linear Regression: 78% KNN: 87%	Early prediction is helpful in predicting the early death cases.	Need to use more datasets.
11	[50]	Real cardiac sounds obtained by the Classifying Heart Sounds Challenge. A smartphone with the iStethoscope Pro software installed was used.	Deep learning model. The 80% data was used of training and 20% data was used for the testing purpose.	The accuracy of 98% was achieved.	It is beneficial to provide the initial level of cardiac diseases screening.	There is a need to evaluate the model on detecting the specific cardiac pathologies. Other sounds related to heart disease can help in understanding the disease.
12	[51]	UCI – Cleveland dataset. 14 attributes were considered out of 76.	LR, KNN, SVM, NB, DT, RF, ANN, MLP and deep neural network	The accuracy obtained for each model was: LR: 93.40%, NB: 90.10%, SVM: 92.30% KNN: 71.42% DT: 81.31% RF: 95.60%, ANN: 92.30%, Deep neural network: 76.92% and MLP: 75.42%	The risk of heart disease can be analyzed.	Other algorithms of deep learning are required to use.
13	[52]	Cleveland heart disease dataset available at UCI. This dataset has 14 features, eight of which are categorical and six of which are numerical.	SVM, NB, LR, RF, and adaboost classifiers	RF achieved the highest accuracy of 86.6%. NB: 83.55%, SVM: 84.46%, LR: 85.07%, Adaboost: 86.59%		It can only determine whether or not a person has cardiac disease. This technique is unable to determine the severity of cardiac disease.

Sr. No.	Source	Dataset	Methods	Results	Advantage	Limitations
14	[53]	Cleveland Dataset, UCI repository, Kaggle and Publicly available heart disease dataset.	DT, NB, SVM, ANN, RF, etc.	Survey has been performed considering different CVD approaches.	The outcomes can be helpful in the early diagnosis.	

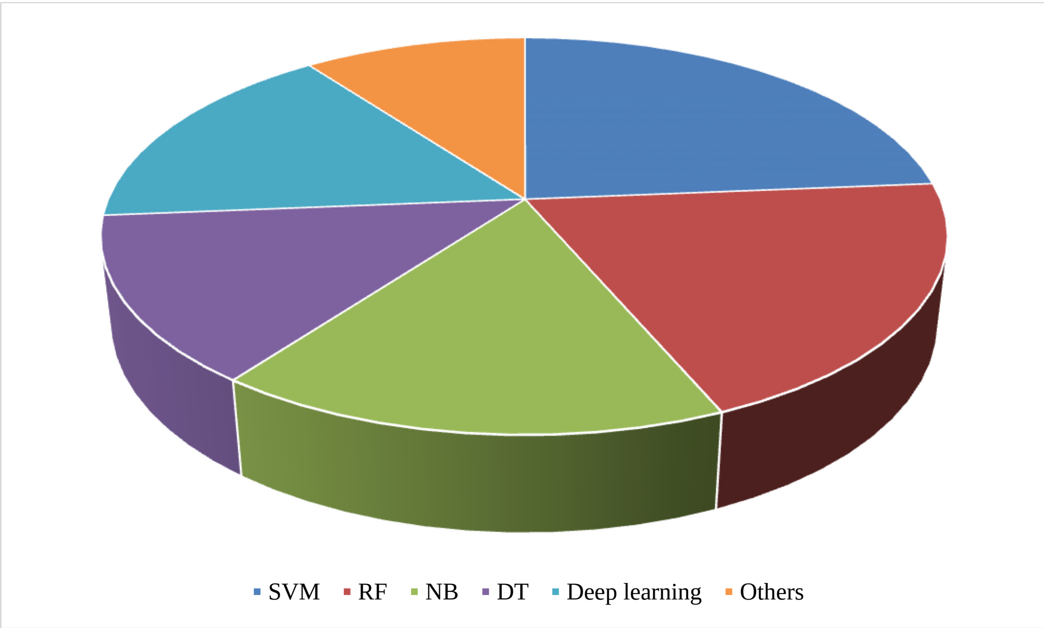

Figure 1 Distribution of the methods based on the applicability in heart disease detection

The related work discussed clearly suggest that the machine learning approaches are useful in the heart disease detection and prediction. Different machine learning approaches are found to be helpful like SVM, RF, ND, DT along with the deep learning and other approaches. It is also found that there is the need of combination of different approaches to improve the performance [54–59].

4.Conclusion

In this paper different methodological aspects have been explored in the direction of heart disease detection and prediction. We have discussed different machine learning approaches like NB, DT, SVM, etc. in the same direction along with other approaches like nature inspired optimization. It is clear from the result that the combination of approaches may be helpful in providing the better results.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

[1] World Health Organization, Cardiovascular Disease (CVDs), 2019. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] American Heart Association. Heart disease and stroke statistics update fact sheet american heart association research heart disease, stroke and other cardiovascular diseases coronary heart disease (CHD). American Heart Association. 2021.

[3] Ramirez-Bautista JA, Hernández-Zavala A, Chaparro-Cárdenas SL, Huerta-Ruelas JA. Review on plantar data analysis for disease diagnosis. Biocybernetics and Biomedical Engineering. 2018; 38(2):342-61.

[4] Nadakinamani RG, Reyana A, Kautish S, Vibith AS, Gupta Y, Abdelwahab SF, et al. Clinical data analysis for prediction of cardiovascular disease using machine

learning techniques. Computational Intelligence and Neuroscience. 2022.
[5] Ansarullah SI, Saif SM, Kumar P, Kirmani MM. Significance of visible non-invasive risk attributes for the initial prediction of heart disease using different machine learning techniques. Computational intelligence and neuroscience. 2022; 2022.
[6] Atallah R, Al-Mousa A. Heart disease detection using machine learning majority voting ensemble method. In 2nd international conference on new trends in computing sciences (ictcs) 2019 (pp. 1-6). IEEE.
[7] Dubey A, Gupta U, Jain S. Medical data clustering and classification using TLBO and machine learning algorithms. Computers, Materials and Continua. 2021; 70(3):4523-43.
[8] Ghosh P, Azam S, Jonkman M, Karim A, Shamrat FJ, Ignatious E, et al. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. IEEE Access. 2021; 9:19304-26.
[9] Patel J, TejalUpadhyay D, Patel S. Heart disease prediction using machine learning and data mining technique. Heart Disease. 2015; 7(1):129-37.
[10] Chahar R, Dubey AK, Narang SK. A review and meta-analysis of machine intelligence approaches for mental health issues and depression detection. International Journal of Advanced Technology and Engineering Exploration. 2021; 8(83):1279-314.
[11] Albahri AS, Zaidan AA, Albahri OS, Zaidan BB, Alamoodi AH, Shareef AH, et al. Development of IoT-based mhealth framework for various cases of heart disease patients. Health and Technology. 2021; 11(5):1013-33.
[12] Dubey AK, Gupta U, Jain S. Comparative study of K-means and fuzzy C-means algorithms on the breast cancer data. International Journal on Advanced Science, Engineering and Information Technology. 2018; 8(1):18-29.
[13] Mehmood A, Iqbal M, Mehmood Z, Irtaza A, Nawaz M, Nazir T, Masood M. Prediction of heart disease using deep convolutional neural networks. Arabian Journal for Science and Engineering. 2021; 46(4):3409-22.
[14] Dubey AK, Narang S, Kumar A, Satya Murthy S, García-Díaz V. Performance estimation of machine learning algorithms in the factor analysis of COVID-19 dataset. Computers, Materials, & Continua. 2021:1921-36.
[15] Arumugam K, Naved M, Shinde PP, Leiva-Chauca O, Huaman-Osorio A, Gonzales-Yanac T. Multiple disease prediction using Machine learning algorithms. Materials Today: Proceedings. 2021.
[16] Xie J, Wu R, Wang H, Chen H, Xu X, Kong Y, Zhang W. Prediction of cardiovascular diseases using weight learning based on density information. Neurocomputing. 2021; 452:566-75.
[17] Gárate-Escamila AK, El Hassani AH, Andrès E. Classification models for heart disease prediction using feature selection and PCA. Informatics in Medicine Unlocked. 2020.
[18] Latha CB, Jeeva SC. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked. 2019.
[19] Chahar R, Dubey AK. A review and analysis of IoT and machine learning algorithms in the brain disease diagnosis and detection. ECS Transactions. 2022; 107(1):6641.
[20] Dubey AK, Kushwaha GR, Shrivastava N. Heterogeneous data mining environment based on dam for mobile computing environments. In international conference on advances in information technology and mobile communication 2011 (pp. 144-9). Springer, Berlin, Heidelberg.
[21] Nemade V, Pathak S, Dubey AK. A systematic literature review of breast cancer diagnosis using machine intelligence techniques. Archives of Computational Methods in Engineering. 2022:1-30.
[22] Dubey AK, Gupta U, Jain S. Analysis of k-means clustering approach on the breast cancer Wisconsin dataset. International Journal of Computer Assisted Radiology and Surgery. 2016; 11(11):2033-47.
[23] Dubey AK, Gupta U, Jain S. Computational measure of cancer using data mining and optimization. In international conference on sustainable communication networks and application 2019 (pp. 626-32). Springer, Cham.
[24] Dubey AK, Kapoor D, Kashyap V. A review on performance analysis of data mining methods in IoT. International Journal of Advanced Technology and Engineering Exploration. 2020; 7(73):193-200.
[25] Tougui I, Jilbab A, El Mhamdi J. Heart disease classification using data mining tools and machine learning techniques. Health and Technology. 2020; 10(5):1137-44.
[26] Kannan R, Vasanthi V. Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease. In soft computing and medical bioinformatics 2019 (pp. 63-72). Springer, Singapore.
[27] Tama BA, Im S, Lee S. Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble. BioMed Research International. 2020.
[28] Saboor A, Usman M, Ali S, Samad A, Abrar MF, Ullah N. A Method for improving prediction of human heart disease using machine learning algorithms. Mobile Information Systems. 2022.
[29] Gonsalves AH, Thabtah F, Mohammad RM, Singh G. Prediction of coronary heart disease using machine learning: an experimental analysis. In proceedings of the 2019 3rd international conference on deep learning technologies 2019 (pp. 51-6).
[30] Phasinam K, Mondal T, Novaliendry D, Yang CH, Dutta C, Shabaz M. Analyzing the performance of machine learning techniques in disease prediction. Journal of Food Quality. 2022.
[31] Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P. Prediction of heart disease using a combination of machine learning and deep learning.

Computational Intelligence and Neuroscience. 2021; 2021.

[32] Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access. 2019; 7:81542-54.

[33] Louridi N, Amar M, El Ouahidi B. Identification of cardiovascular diseases using machine learning. In 7th mediterranean congress of telecommunications (CMT) 2019 (pp. 1-6). IEEE.

[34] Alotaibi FS. Implementation of machine learning model to predict heart failure disease. International Journal of Advanced Computer Science and Applications. 2019; 10(6):261-8.

[35] Rajdhan A, Agarwal A, Sai M, Ravi D, Ghuli P. Heart disease prediction using machine learning. International Journal of Research and Technology. 2020; 9(04):659-62.

[36] Kavitha M, Gnaneswar G, Dinesh R, Sai YR, Suraj RS. Heart disease prediction using hybrid machine learning model. In 6th international conference on inventive computation technologies (ICICT) 2021 (pp. 1329-33). IEEE.

[37] Sharma P, Choudhary K, Gupta K, Chawla R, Gupta D, Sharma A. Artificial plant optimization algorithm to detect heart rate & presence of heart disease using machine learning. Artificial Intelligence in Medicine. 2020; 102:101752.

[38] Mezzatesta S, Torino C, De Meo P, Fiumara G, Vilasi A. A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis. Computer Methods and Programs in Biomedicine. 2019; 177:9-15.

[39] Ali F, El-Sappagh S, Islam SR, Kwak D, Ali A, Imran M, Kwak KS. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. Information Fusion. 2020; 63:208-22.

[40] Tarawneh M, Embarak O. Hybrid approach for heart disease prediction using data mining techniques. In international conference on emerging internetworking, data & web technologies 2019 (pp. 447-54). Springer, Cham.

[41] Maji S, Arora S. Decision tree algorithms for prediction of heart disease. In information and communication technology for competitive strategies 2019 (pp. 447-54). Springer, Singapore.

[42] Khourdifi Y, Bahaj M. Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. International Journal of Intelligent Engineering and Systems. 2019; 12(1):242-52.

[43] Repaka AN, Ravikanti SD, Franklin RG. Design and implementing heart disease prediction using naives Bayesian. In international conference on trends in electronics and informatics (ICOEI) 2019 (pp. 292-7). IEEE.

[44] Ganesan M, Sivakumar N. IoT based heart disease prediction and diagnosis model for healthcare using machine learning models. In international conference on system, computation, automation and networking (ICSCAN) 2019 (pp. 1-5). IEEE.

[45] Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Medical Informatics and Decision Making. 2019; 19(1):1-5.

[46] Krittanawong C, Virk HU, Bangalore S, Wang Z, Johnson KW, Pinotti R, et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. Scientific Reports. 2020; 10(1):1-1.

[47] Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. SN Computer Science. 2020; 1(6):1-6.

[48] Khan MA. An IoT framework for heart disease prediction based on MDCNN classifier. IEEE Access. 2020; 8:34717-27.

[49] Singh A, Kumar R. Heart disease prediction using machine learning algorithms. In international conference on electrical and electronics engineering (ICE3) 2020 (pp. 452-7). IEEE.

[50] Brunese L, Martinelli F, Mercaldo F, Santone A. Deep learning for heart disease detection through cardiac sounds. Procedia Computer Science. 2020; 176:2202-11.

[51] Katarya R, Meena SK. Machine learning techniques for heart disease prediction: a comparative study and analysis. Health and Technology. 2021; 11(1):87-97.

[52] Rani P, Kumar R, Ahmed NM, Jain A. A decision support system for heart disease prediction based upon machine learning. Journal of Reliable Intelligent Environments. 2021; 7(3):263-75.

[53] Swathy M, Saruladha K. A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deep learning techniques. ICT Express. 2022; 8(1):109-16.

[54] Dubey AK, Choudhary K. A systematic review and analysis of the heart disease prediction methodology. International Journal of Advanced Computer Research. 2018; 8(38):240-56.

[55] Dubey AK, Choudhary K, Sharma R. Predicting heart disease based on influential features with machine learning. Intelligent Automation and Soft Computing. 2021; 30(3):929-43.

[56] Dubey AK, Sinhal AK, Sharma R. An improved auto categorical PSO with ML for heart disease prediction. Engineering, Technology & Applied Science Research. 2022; 12(3):8567-73.

[57] Dubey A, Patel R, Choure K. An efficient data mining and ant colony optimization technique (DMACO) for heart disease prediction. International Journal of Advanced Technology and Engineering Exploration. 2014; 1(1):1-6.

[58] Ahsan MM, Siddique Z. Machine learning-based heart disease diagnosis: a systematic literature review. Artificial Intelligence in Medicine. 2022: 102289.

[59] El-Hasnony IM, Elzeki OM, Alshehri A, Salem H. Multi-label active learning-based machine learning model for heart disease prediction. Sensors. 2022; 22(3):1184.

Animesh Kumar Dubey is working as Assistant professor with the department of Computer Science and Engineering, at Madhyanchal Professional University, Bhopal, India. He has completed his Bachelor of Engineering (B.E.) and MTech. degree with Computer Science Engineering from Rajeev Gandhi Technical University, Bhopal (M.P.). He has more than 15 publications in reputed, peer-reviewed national and international journals and conferences. His research areas are Data Mining, Optimization, Machine Learning, Cloud Computing and Artificial Intelligence.
Email:animeshdubey123@gmail.com

Dr. Mohd Zuber is working as Associate professor with the department of Computer Science and Engineering, at Madhyanchal Professional University, Bhopal, India. He has completed his P.hD. degree with Computer Science from Ravindra Nath Tegore University, Bhopal (M.P.). He has more than 7 publications in reputed, peer-reviewed national and international journals and conferences. He is a Memeber of 4 international or national Research journals and also reviwers. His research areas are Data Mining,Cyber security, Machine Learning, Cloud Computing and Artificial Intelligence.
Email: mzmkhanugc@gmail.com