International Conference on Innovative Data Communication Technologies and Application (ICIDCA-2023)
IEEE Xplore Part Number: CFP23CR5-ART; ISBN: 979-8-3503-9720-8

Improving Accuracy of Heart Disease Prediction through Machine Learning Algorithms

Mohammed Ali Shaik
School of Computer Science &Artificial Intelligence
SR University
Warangal, Telangana State, India
niharali@gmail.com

Radhandi Sreeja
Department of Computer Science & Engineering
SR University
Warangal, Telangana State, India
sreejaradhandi@gmail.com

Safa Zainab
Department of Computer Science & Engineering
SR University
Warangal, Telangana State, India
zainabsafa310@gmail.com

Panthangi Shiva Sowmya
Department of Computer Science & Engineering
SR University
Warangal, Telangana State, India
sowmyapanthangi555@gmail.com

Thipparthy Akshay
Department of Computer Science & Engineering
SR University
Warangal, Telangana State, India
akshaythipparthy1@gmail.com

Sudireddy Sindhu
Department of Computer Science & Engineering
SR University
Warangal, Telangana State, India
sindhusudireddy38@gmail.com

2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA) | 979-8-3503-9720-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIDCA56705.2023.10100244

***Abstract*--One of the main causes of death in the world is heart disease. Early heart disease detection and treatment can lower mortality rates and enhance quality of life which is the major challenge. "Machine learning" algorithms can accurately predict the likelihood of getting the heart disease by using data like: "age, gender, lifestyle factors, medical history, and laboratory testing". Building a ML model for "heart disease prediction" which is merely relies on the various relevant factors is the primary goal of this paper. For this research project, we used 4 different datasets which comprises of distinct factors that are relevant to heart disease. The model building is made through ML algorithms: "Random forest, K-nearest neighbour, logistic regression, and decision tree". The study demonstrates that, when compared to other ML techniques, logistic regression and KNN provide better prediction accuracy in a shorter amount of time.**

Keywords — Heart Disease, K-nearest neighbour, Logistic regression, Decision Tree, Random forest, data pre-processing, Classification, Machine learning , Accuracy.

I. INTRODUCTION

In accordance with the World Health Organization [3], peripheral arterial disease and coronary heart disease are all heart and blood vessel issues that are categorised as cardiovascular diseases (CVDs). Cardiovascular disorders caused 17.9 million deaths globally in 2016, representing for 31% of all deaths. 85% of these deaths were caused by heart attacks and strokes. More than 75% of heart-related deaths happen in low- and middle-income countries, one of the main risk factors for acquiring heart diseases is high blood pressure.. In 2016, 63% of all NCD-related deaths occurred in India, with cardiovascular disease accounting for 27% of those fatalities. Moreover, cardiovascular disease accounts for 45% of deaths in persons between the ages of 40 and 69 [1]. The primary blood arteries that supply the heart muscle are affected by a disease called as coronary artery disease. Heart disease is typically brought on by plaques, which are thickened cholesterol clots in the heart arteries. Atherosclerosis describes the clogging of these arteries. The heart as well as other body organs receive less blood as a result of atherosclerosis. It may result in a heart attack, angina, or a stroke [2].

With this in mind, software that facilitates the advance detection of heart disease has indeed been created using computing technology and "Machine Learning (ML)" techniques[12]. Any of the heart condition that is detected early can have a positive impact on mortality. In order to comprehend data patterns and create predictions based on them, many machine learning approaches are applied in the medicine field. Usually, "healthcare" data is very large in both volume and structure. Large data sets may be handled by most of the ML algorithms, which can then be used for mining the information as ML algorithms will make predictions about the future by learning from the past. Cardiologists may be inspired to act more rapidly to treat more patients by employing this kind of ML framework for coronary disease prediction [14].

Extreme amounts of complicated data are being put into healthcare settings via labs, medical devices, physicians' notes, and other sources. Although electronic health records are assisting in the digitization of the data, they are not intended to reduce the administrative burden or facilitate quick decisions. The value of all the incoming data depends on how soon insights can be drawn from it and put to use to enhance healthcare delivery. That is achievable thanks to machine learning, especially for digital data sets with recognisable patterns. Data from many sources is not only collected by machine learning, but it is also unified. It can carry out the complex computations needed by physicians, nurses, and other healthcare professionals to quickly interpret unprocessed physiological, behavioural, and imaging data [3].

II. LITERATURE REVIEW

A. Feature Extraction

By creating new features from old ones, feature extraction aims to reduce the amount of features that are available in a dataset as the majority of the data contained in the original set of characteristics that are capable of summarizing the correspondingly lower collection of features [4]. Utilizing feature extraction techniques also has the following advantages: increased precision, reducing the chance of overfitting, increase training speed, better data visualization, improve our model's explainability [15]. And the major ML based feature extraction techniques are

- Principal Component Analysis (PCA): A dimensionality reduction technique that helps in the extraction of the most important features from a dataset. It reduces the number of parameters in the dataset while retaining as much variance as possible [1].
- Linear Discriminant Analysis (LDA): identifies the ideal linear combination of characteristics that distinguishes between two or more classes of events or objects using supervised learning. In order to maintain even more class separation as possible, it is used to reduce a dataset[2].
- Feature Selection: is a vital step in ML and data mining processes that removes irrelevant or redundant features from the dataset before further processing [9]. This improves accuracy while also lowering the computational costs associated with training models on large datasets[32].
- Feature Extraction: Feature extraction is the process of changing original data into more meaningful representations, such as main components or other numerical representations that capture key information about the data[23]. When training models on huge datasets, noise is reduced and accuracy is improved [3].
- Clustering: Clustering is an unsupervised learning technique that groups together similar objects based on their attributes or features [24]. This can be used to identify data patterns and extract useful information from them [5].

B. Datasets

We used four different datasets to create this model. They are the UCI Heart Disease dataset, the Cleveland Heart Disease dataset, the Framingham Heart Study dataset, and the Cardiovascular Disease dataset [20].

1. UCI Heart Disease Dataset

The UCI Heart Disease dataset is a collection of medical records from patients who have been diagnosed with heart disease as the dataset contains 303 records, each with 14 attributes, including age, sex, chest pain type, resting blood pressure, serum cholesterol levels, and maximum heart rate achieved[29]. The goal of the dataset is to perform prediction of the existence or nonexistence of heart disease in a given patient based on these attributes. This dataset has been widely used in research and has been used to develop various machine learning models for predicting heart disease [31].

2. Framingham Heart Study dataset

The Framingham Heart Study is a lengthy, continual cardiovascular cohort study of people living in the Massachusetts town of Framingham. With 5,209 adult Framingham volunteers at the study's outset, it has now recruited its third generation of participation. Finding common elements or traits that lead to cardiovascular disease is the study's main goal [30]. Data collected includes demographic information, lifestyle habits, medical history, and laboratory tests. The data has been used to identify risk factors for heart disease and stroke, as well as to develop guidelines for prevention and treatment. Over 4,240 records, 16 columns, and 15 attributes comprise the "Framingham" heart disease dataset [32].

3. Heart Disease Cleveland UCI dataset

The “Cleveland Heart Disease dataset” is a collection of medical records from the Cleveland Clinic Foundation. It contains 14 attributes, including patient age, sex, blood pressure, cholesterol levels, and other medical attributes. The dataset also includes a target variable that indicates whether or not the patient has “heart disease” or not [8].

4. Cardiovascular Disease dataset

This dataset is a collection of data related to the diagnosis and treatment of cardiovascular diseases. It includes information on patient demographics, medical history, laboratory tests, imaging studies, and other clinical data[31]. The dataset can be used to identify risk factors for cardiovascular disease and to develop predictive models for early detection and prevention. It can also be used to study the effectiveness of treatments and to compare outcomes between different patient populations [9].

C. Machine Learning Algorithms

This model was created using four ML techniques are:

1. Logistic Regression (LR)

LR is a classification which is task based supervised ML technique that predicts the likelihood of a categorical dependent variable based on one or more independent variables as the dependent variable in logistic regression is binary, which implies it has just two possible outcomes (e.g., success or failure, yes or no, 0 or 1) [16]. Using an equation, logistic regression calculates the chance that a given data point belongs to one of two categories [25]. The equation considers the independent variables and assigns each a weight depending on their importance in determining the result. The model then generates predictions based on these scores and probabilities [26]. The logistic regression score must fall within the range of 0 and 1, and it cannot exceeds this limit, as it produces a "S"-shaped curve which is also known as the “logistic function or sigmoid function” [10].

2. Random Forest (RF)

RF is a popular supervised ML method which can be used in ML built on the idea of ensemble learning, which combines difficult problems to be tackled in order to resolve classification and regression problems and enhances the performance aspects of the model[17]. According to its name, “Random Forest is indeed a classifier typically includes several decision trees on various subsets of a provided dataset that calculates the average to boost the projected accuracy of that dataset” [14]. The RF collects projections from every tree and forecasts the ultimate results that are based on the clear majority of predictions being opposed to rely over just one decision tree developed [11].

3. Decision Tree

The idea of ensemble learning, which conducts numerous classifications by integrating challenging issues that must be resolved and improve the model's performance in order to tackle classification and regression problems [22]. RF is a classifier that, as the name suggests to combine a variety of decision trees over distinct subsets for evaluating the dataset over its averages attained to improve the projected accuracy rate of that dataset [9]. Instead than relying on a single decision tree, the RF collects predictions from each tree and forecasts the ultimate result using the predictions that received the most votes [12].

4. K-Nearest Neighbour (KNN)

A machine learning method called K-Nearest Neighbor uses the supervised learning approach [21].The KNN algorithm allocates the test instance to the classification which is most comparable to the previous grouping on the assumption that the new model and the prior examples are comparable [19]. Subsequent to all the previous data has been saved, a fresh collection of data is classified using the KNN algorithm which is merely based on similarity aspects [18]. This means that fresh data may be rapidly and properly categorized using the KNN approach which is most commonly used to solve classification issues , it may also be used to solve regression difficulties [13].

D. *Challenges and Limitations*

The challenges are:

- Accurately predicting heart disease is difficult due to the complexity of the disease and the many factors that can contribute to it.
- Gathering accurate and comprehensive data about a patient's medical history and lifestyle habits can be difficult.
- Identifying subtle changes in a patient's health that could indicate an increased risk of heart disease can be challenging [27].
- Developing effective predictive models requires a large amount of data and computing power, which can be expensive and time consuming to obtain.

The Limitations are:

- Predictive models are limited by the accuracy of the data used to create them, so any errors or inaccuracies in the data can lead to inaccurate predictions.
- Predictive models are also limited by their ability to accurately predict future outcomes based on past data, so they may not always be able to accurately predict future events or trends in a patient's health or lifestyle habits.
- Predictive models are also limited by their ability to identify subtle changes in a patient's health that could indicate an increased risk of heart disease, as these changes may not always be obvious or easily detectable with current technology or methods [28].

III. Proposed Methodology

Attribute selection is performed through decision tree learning approach and random forest learning approach.

A. *Data Collection:* We used four different datasets to create this model. They are the "UCI Heart Disease dataset", the "Cleveland Heart Disease dataset", the "Framingham Heart Study dataset", and the "Cardiovascular Disease dataset".

B. *Data Preprocessing*: Pre-processing data is a method for transforming unclean data into a clean dataset. The data was collected from several sources in raw format, making analysis impractical.

- Attribute selection: Only few attributes are selected to train the model inorder to get best accuracy and to delete noisy data or irrelevent data.
- missing value treatment: Missing value treatment is implemented to remove the null or missing values or fill them with the mean values. This step is performed if there are any missing values in dataset.
- Dummy values treatment: This step is implemented if the data contains categorical values.
- Normalization: Normalizing the values using "min-max scaler". (optional)

C. *Splitting data*: The pre-processed data is then divided into 2parts in the ratio 90:10.They are train data and test data. The train data is used for constructing the model. The test data is used to test the accuracy.

D. *Model Construction*: After splitting the data , the model is constructed using some machine learning algorithms such as SVC, Random Forest Classifier. The data is trained using the algorithms and calculate the accuracies for each algorithm. At last, one model is finalized which is having good accuracy.

E. *Testing*: After training the model, We need to test our model with the help of testing data and calculate the accuracy and error metrics.

F. *Web application*: With the help of flask, we are going to create a web application.

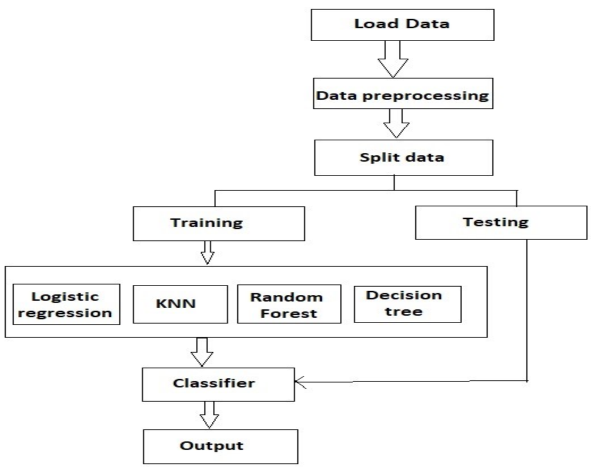

Fig. 1. Flow chart of the proposed model

Figure1 depicts the workflow of the proposed model. It includes series of steps to construct the model. It also includes algorithms that were used.

Algorithm	Accuracy		Error Metrics		
	Training	Testing	MSE	RMSE	MAE
Logistic Regression	64.93	64.39	**0.35**	**0.59**	**0.59**
KNN	8.79	58.9	0.41	0.64	0.41
Random Forest	73.27	72.74	0.27	0.52	0.27
Decision Tree	**73.45**	**73.04**	0.26	0.514	0.26

IV. RESULTS

Algorithm	Accuracy		Error Metrics		
	Training	Testing	MSE	RMSE	MAE
Logistic Regression	85.78	**84.59**	0.15	0.392	0.15
KNN	**88.58**	80.97	**0.19**	**0.43**	0.18
Random Forest	85	83	0.154	0.394	0.154
Decision Tree	86.89	82.63	0.17	0.416	**0.486**

After evaluating the proposed methodology the results obtained are depicted in terms of tables and figures as follows

Table 1: Accuracy and error metrics using “heart.csv” dataset from UCI Dataset Repository.
Table 2: Accuracy and error metrics using “heart_cleveland_upload.csv” dataset collected from Kaggle.
Table 3: Accuracy and error metrics using “cardio.csv” (cardiovascular) dataset collected from Kaggle.

Table 4: accuracy and error metrics using “framingham.csv” dataset collected from “Kaggle”

The training and testing accuracy percentages for four algorithms employing data sourced from the "UCI dataset repository" are shown in Table 1. In this situation, Logistic Regression works well and accurately predicts the outcome 93.55% of the time. A decision tree can predict with at least 74% accuracy.

Using the “heart_Cleveland_upload.csv” dataset obtained from "Kaggle," Table 2 shows the training and testing accuracy percentages for four methods. KNN performs well in this situation and accurately predicts the outcome 90% of the time. Decision tree forecasts with a 76.67% accuracy rate. Additionally, mean square and room mean square error statistics are included.

The training and testing accuracy percentages for four algorithms utilising the “cardio.csv” dataset obtained from "Kaggle" are shown in Table 3. In this case, decision tree forecasts the outcome accurately (73.04%). Additionally, mean square and room mean square error statistics are included.

The training and testing accuracy percentages for four algorithms utilising the “framingham.csv” dataset obtained from “Kaggle” are shown in Table 4. In this case, logistic regression performs well and accurately predicts the outcome (84.59%). Additionally, mean square and room mean square error statistics are included.

And the results obtained are:

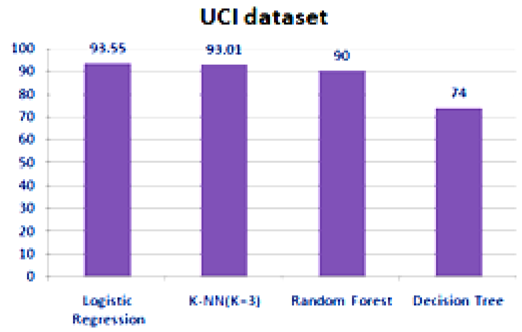

Fig. 2. Four algorithms' accuracy rates for “UCI” dataset

The graphical depiction of the accuracy percentages of the four employed algorithms is shown in Figure 2. The best method is logistic regression, which has a 93.55% accuracy rate. Decision trees have a prediction accuracy of at least 74%.

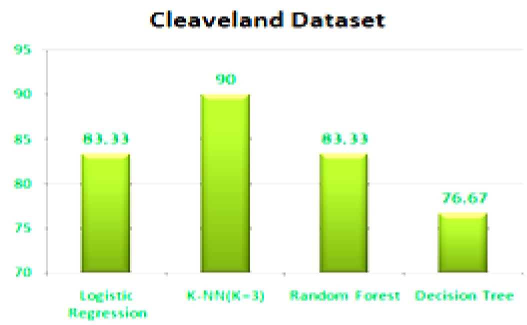

Fig. 3. Four algorithms' accuracy rates for “Cleaveland” dataset

The graphical depiction of the accuracy percentages for the four employed methods is shown in Figure 3. KNN performed well, achieving 90% accuracy. The least accurate prediction made by a decision tree is 76.67%.

Algorithm	**Accuracy**		**Error Metrics**		
	Training	**Testing**	**MSE**	**RMSE**	**MAE**
Logistic Regression	85.29	**93.55**	0.064	0.25	0.06
K-NN(K=3)	88	93.01	0.065	0.254	0.06
Random Forest	93	90	**0.096**	0.311	0.096
Decision Tree	92	74	0.258	**0.508**	**0.258**

Algorithm	**Accuracy**		**Error Metrics**		
	Training	**Testing**	**MSE**	**RMSE**	**MAE**
Logistic Regression	85.77	83.33	0.166	0.25	0.16
KNN	89.89	90	0.1	0.316	0.1
Random Forest	93.26	83.33	0.16	0.408	0.16
Decision Tree	92.88	76.67	0.23	0.483	0.486

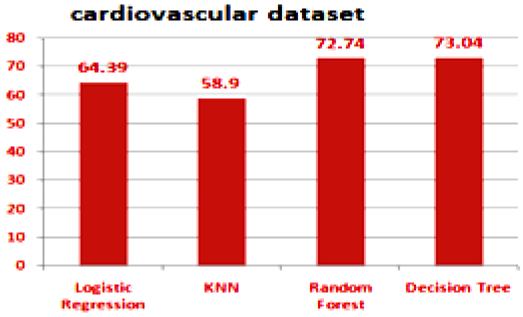

Fig. 4. Four algorithms' accuracy rates for "Cardiovascular" dataset

The graphical depiction of the accuracy percentages for the four employed methods is shown in Figure 4. The accuracy of the decision tree was 73.04%, whereas the accuracy of the random forest was 72.24%.

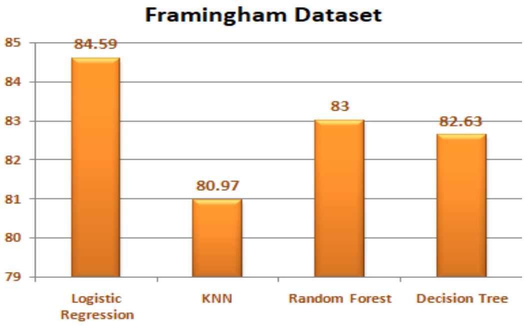

Fig. 5. Four algorithms' accuracy rates for "Framingham" dataset

The model training, Figure 5 shows the accuracy rates of four algorithms for the four datasets employed. We can see that the dataset used to do the logistic regression has an accuracy of 84.59%.

V. CONCLUSION

After implementing a ML approach for performing training and testing, we found that the accuracy of the Logistic Regression is significantly more effective than other methods. Each algorithm's confusion matrix, error metrics, and accuracy score are used to evaluate performance. We achieved a 93.55% accuracy using logistic regression using data that was taken from the UCI repository.KNN likewise predicts well, with an accuracy of 93.01%.In the future, heart disease prediction could be improved by incorporating more data sources such as genetics, lifestyle, and environmental factors. Machine learning algorithms could be used to identify patterns in the data and predict the risk of heart disease. Additionally, artificial intelligence (AI) could be used to better understand the complex relationships between different risk factors and their impact on heart health. AI could also be used to develop personalized treatments for individuals based on their individual risk profiles. Finally, wearable technology could be used to monitor vital signs and provide early warning signs of potential heart problems.

REFERENCES

[1] M. J. A. Junaid and R. Kumar, "Data Science And Its Application In Heart Disease Prediction," 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 2020, pp. 396-400, doi: 10.1109/ICIEM48762.2020.9160056.

[2] K. S. K. Reddy and K. V. Kanimozhi, "Novel Intelligent Model for Heart Disease Prediction using Dynamic KNN (DKNN) with improved accuracy over SVM," 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2022, pp. 1-5, doi: 10.1109/ICBATS54253.2022.9758996.

[3] G. S. Reddy Thummala and R. Baskar, "Prediction of Heart Disease using Decision Tree in Comparison with KNN to Improve Accuracy," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2022, pp. 1-5, doi: 10.1109/ICSES55317.2022.9914044.

[4] T. P. Naidu et al., "A Hybridized Model for the Prediction of Heart Disease using ML Algorithms," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2021, pp. 256-261, doi: 10.1109/ICAC3N53548.2021.9725780.

[5] Purushottam, K. Saxena and R. Sharma, "Efficient heart disease prediction system using decision tree," International Conference on Computing, Communication & Automation, Greater Noida, India, 2015, pp. 72-77, doi: 10.1109/CCAA.2015.7148346.

[6] M. T. Islam, S. R. Rafa and M. G. Kibria, "Early Prediction of Heart Disease Using PCA and Hybrid Genetic Algorithm with k-Means," 2020 23rd International Conference on Computer and Information Technology (ICCIT), DHAKA, Bangladesh, 2020, pp. 1-6, doi: 10.1109/ICCIT51783.2020.9392655.

[7] S. K. L. Sameer and P. Sriramya, "Improving the Accuracy for Prediction of Heart Disease by Novel Feature Selection Scheme using Decision tree comparing with Naive-Bayes Classifier Algorithms," 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2022, pp. 1-8, doi: 10.1109/ICBATS54253.2022.9758926.

[8] D. D. Saikumar, U. P. P. S and I. M. Arumugam, "Prediction of Heart Disease Using K-Nearest Neighbour Algorithm in Comparison with Support Vector Machine Algorithm," 2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), Karachi, Pakistan, 2022, pp. 1-8, doi: 10.1109/MACS56771.2022.10023034.

[9] Y. Liu, M. Zhang, Z. Fan and Y. Chen, "Heart disease prediction based on random forest and LSTM," 2020 2nd International Conference on Information Technology and Computer Application (ITCA), Guangzhou, China, 2020, pp. 630-635, doi: 10.1109/ITCA52113.2020.00137.

[10] Mohammed Ali Shaik, Praveen Pappula, T Sampath Kumar, Predicting Hypothyroid Disease using Ensemble Models through Machine Learning Approach, European Journal of Molecular & Clinical Medicine, , 9(7), 6738-6745, (2022).

[11] M. A. Shaik, S. k. Koppula, M. Rafiuddin and B. S. Preethi, COVID-19 Detector Using Deep Learning, International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 443-449, (2022).

[12] A. Basak, M. S. Rahman and M. Rahman, "Prediction of Heart Disease Using an Approach Based on Machine Learning," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-5, doi: 10.1109/ICCCNT54827.2022.9984555.

[13] Mohammed Ali Shaik and Dhanraj Verma, Prediction of Heart Disease using Swarm Intelligence based Machine Learning Algorithms, International Conference on Research in Sciences, Engineering & Technology, Published by AIP Publishing. 978-0-7354-4368-6, 2022, pp. 020025-1 to 020025-9.

[14] X. Huang, H. Fan, H. Zhu and X. Zhang, "Prediction of Heart Disease based on Enhanced Random Forest," 2021 IEEE 7th International Conference on Cloud Computing and Intelligent Systems (CCIS), Xi'an, China, 2021, pp. 398-402, doi: 10.1109/CCIS53392.2021.9754669.

[15] A. Kumari and A. K. Mehta, "A Novel Approach for Prediction of Heart Disease using Machine Learning Algorithms," 2021 Asian Conference on Innovation in Technology (ASIANCON), PUNE, India, 2021, pp. 1-5, doi: 10.1109/ASIANCON51346.2021.9544544.

[16] C. Sowmiya and P. Sumitra, "Analytical study of heart disease diagnosis using classification techniques," 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Srivilliputtur, India, 2017, pp. 1-5, doi: 10.1109/ITCOSP.2017.8303115.

[17] Mohammed Ali Shaik, Geetha Manoharan, B Prashanth, NuneAkhil, Anumandla Akash and Thudi Raja Shekhar Reddy, "Prediction of Crop Yield using Machine Learning", International Conference on Research in Sciences, Engineering & Technology, AIP Conf. Proc. 2418, (2022),pp. 020072-1–020072-8,

[18] Mohammed Ali Shaik and Dhanraj Verma, Enhanced ANN training model to smooth and time series forecast, IOP Conf. Ser.: Mater. Sci. Eng. 981, (2020), pp. 022038

[19] A. M. Tushar, A. Wazed, E. Shawon, M. Rahman, M. I. Hossen and M. Z. H. Jesmeen, "A Review of Commonly used Machine Learning Classifiers in Heart Disease Prediction," 2022 IEEE 10th Conference on Systems, Process & Control (ICSPC), Malacca, Malaysia, 2022, pp. 319-323, doi: 10.1109/ICSPC55597.2022.10001742.

[20] Mohammed Ali Shaik, T. Sampath Kumar, P. Praveen, R. Vijayaprakash, Research on Multi-Agent Experiment in Clustering", International Journal of Recent Technology and Engineering (IJRTE), 8(1S4), 1126-1129,(2019).

[21] L. P. Koyi, T. Borra and G. L. V. Prasad, "A Research Survey on State of the art Heart Disease Prediction Systems," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 799-806, doi: 10.1109/ICAIS50930.2021.9395785.

[22] Mohammed Ali Shaik and Dhanraj Verma, Deep learning time series to forecast COVID-19 active cases in INDIA: A comparative study, IOP Conf. Ser.:Mater.Sci.Eng. 981,(2020), pp. 022041.

[23] P Praveen, M Ranjith Kumar, Mohammed Ali Shaik, R Ravi kumar and R Kiran, The comparative study on agglomerative hierarchical clustering using numerical data, IOP Conf. Ser.: Mater. Sci. Eng. 981, 2020, pp. 022071.

[24] W. A. Husna Wan Azizan, A. Afhzan Ab Rahim, S. L. Mohd Hassan, I. S. Abdul Halim and N. E. Abdullah, "A Comparative Study of Two Machine Learning Algorithms for Heart Disease Prediction System," 2021 IEEE 12th Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Malaysia, 2021, pp. 132-137, doi: 10.1109/ICSGRC53186.2021.9515250.

[25] D Kothandaraman, N Praveena, K Varadarajkumar, B Madhav Rao, Dharmesh Dhabliya, Shivaprasad Satla, Worku Abera, Intelligent Forecasting of Air Quality and Pollution Prediction Using Machine Learning, Adsorption Science & Technology, 2022, (2022).

[26] Sallauddin Mohmmad, Ramesh Dadi, D Kothandaraman, E Sudarshan, Syed Nawaz Pasha, Mohammed Ali Shaik, A survey machine learning based object detections in an image, AIP Conference Proceedings,2418, (1), (2022), pp. 020024.

[27] A Balasundaram, D Kothandaraman, S Ashokkumar, E Sudarshan, Chest X-ray image based COVID prediction using machine learning, AIP Conference Proceedings, 2418(1), 2022, pp. 020079.

[28] Mohammed Ali Shaik, Time Series Forecasting using Vector quantization", International Journal of Advanced Science and Technology (IJAST), 29(4), 169-175,(2020).

[29] Mohammed Ali Shaik, A Survey on Text Classification methods through Machine Learning Methods", International Journalof Control and Automation (IJCA),12(6), 390-396, (2019).

[30] R. Ravi Kumar, M. Babu Reddy and P. Praveen, A review of feature subset selection on unsupervised learning Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), (2017), pp.163-167.

[31] UCI Machine Learning Repository: Heart Disease Data Set,Link: https://archive.ics.uci.edu/ml/datasets/Heart+Disease

[32] https://www.framinghamheartstudy.org/fhs-overview/about-the-stud/

[33] https://archive.ics.uci.edu/ml/datasets/Heart+Disease.