xx

COPY RIGHT

Paper Authors **Vaishnavi Shetty Mogilipally, Ajmeera Preethi, G Venkanna**

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER
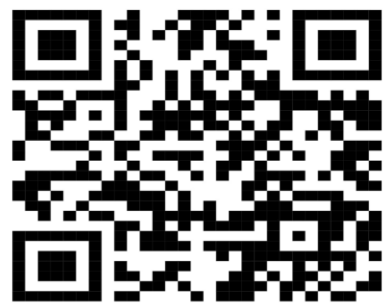
To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

ENSEMBLE MODEL FOR EXPLORATORY DATA ANALYSIS AND PREDICTION OF CARDIOMYOPATHY

1 . Vaishnavi Shetty Mogilipally, 2.Ajmeera Preethi, 3.G Venkanna

1 2 Students , SNIST(Sreenidhi Institute of science and technology).

3 Asst.Professor , SNIST(Sreenidhi Institute of science and technology) venkannag@sreenidhi.edu.in

Abstract: Cardiomyopathy is a dangerous disease of the heart muscles that can cause heart failure and other problems that can be life-threatening. It refers to a group of heart conditions that change the heart's structure and make it harder for it to work. It can be caused by genes, diseases, or things in the surroundings. Cardiomyopathy must be found early on so that therapy can begin on time and patient results are improved. Traditional ways of diagnosing diseases might not be sensitive enough to pick up on small signs of illness, which would mean that treatment would have to be put off. So, this paper suggests a new way of doing things that is called the Ensemble Model for Exploratory Data Analysis and Prediction of Cardiomyopathy (EMEDAPC). The suggested EMEDAPC offers a complete way to look at various kinds of statistics about cardiomyopathy. This model can accurately find important traits linked to the disease, which leads to better prediction. It also greatly lowers the chance of fake positives and blanks, giving more accurate results for diagnosing cardiomyopathy and figuring out the risk of it. The study could change cardiology by giving doctors more information to help them make better choices, allowing early treatments, and eventually making the lives of people with cardiomyopathy better.

Index terms – Cardiomyopathy, Exploratory Data Analysis, Prediction.

1. INTRODUCTION

CVD kills one American every 36 seconds. It kills over 665 million people—one in four. Heart disease and stroke are expensive for US healthcare. Medicine, medical treatment, and death-related lost work time cost $219 billion in 2014 and 2015. Early treatment may prevent fatal heart failure. Many believe angiography is the greatest approach to anticipate coronary artery disease (CAD), yet it's costly and difficult for low-income households to afford.

Diabetes, high blood pressure, high cholesterol, irregular heartbeat, and other risk factors make heart disease hard to detect. Heart disease involves the heart and veins. Different types of heart disease have different symptoms. A competent and experienced person can find and diagnose arterial disease.

Age, diabetes, smoking, obesity, junk food, and other factors contribute. Several factors may cause or worsen heart disease. Management software monitors patient and professional data at most hospitals. It's popular, and these instruments generate plenty of patient data. These information seldom influence physicians' judgments. These data are valuable and

hold a lot of unused information. Much clinical data has to be transformed into useful information to allow smart technologies aid healthcare practitioners make choices. Medical data processing was studied since there weren't enough professionals and many instances were misidentified. A fast, precise automated detecting system was required. The major objective is to utilize the classifier model to categorize crucial medical data and produce early heart disease predictions.

WHO claims heart illnesses kill most people worldwide. Many factors may cause cardiovascular disease (CVD), including excessive blood pressure, cholesterol, diabetes, irregular heartbeat, and more. This makes CVD diagnosis difficult. Men and women may have distinct CVD symptoms. Men are more likely to experience chest discomfort, while women are more likely to have other symptoms including feeling unwell, being exhausted, and having difficulties breathing. Researchers have been looking into a lot of different ways to predict heart diseases. However, predicting diseases early on is not very useful because of a number of factors, such as the approach's complexity, execution time, and accuracy. So, getting the right care and evaluation can save a lot of lives.

The heart muscle has problems with its structure and function, which make cardiomyopathy a complicated and diverse medical disease. Cardiomyopathy is a major public health issue around the world. It can cause problems with the heart that can kill you, like heart failure and irregular heartbeats. It is very important to understand the reasons, processes, and possible solutions for cardiomyopathy in order to improve the health and well-being of people who have it and make healthcare systems less busy.

Cardiomyopathy can be caused by many things, such as changes in genes, infections, autoimmune diseases, metabolic disorders, and being exposed to drugs or toxins. Researchers are interested in these different causes because they want to create methods and treatments that are more effective based on the underlying factors. Some types of cardiomyopathy, like hypertrophic cardiomyopathy and enlarged cardiomyopathy, are strongly linked to genes. Finding the DNA changes that cause these conditions could help with figuring out the risks, diagnosing them early, and maybe even coming up with gene-based treatments. We can better identify and keep an eye on cardiomyopathy thanks to progress in medical imaging, genetics, and molecular biology. Researchers want to see how these tools can be used to improve early identification and individual treatment plans.

One of the main goals is to improve the quality of life for people who have cardiomyopathy. Not only does this kind of research look into the physical effects of the disease, but it also looks into the mental and emotional effects it has on patients and their families. It is very important to know what causes cardiomyopathy and how to avoid getting it. Cardiomyopathy risk can be lowered by making changes to your lifestyle, like what you eat and how much you exercise, as well as by taking medicine. Researchers want to find successful ways to stop this from happening.

Cardiomyopathy affects people of all ages and walks of life around the world. A big reason why people are researching cardiomyopathy is because of the work of

groups that help patients and caregivers. These groups help people who have the disease by spreading information about it, giving them support, and funding study projects that aim to make their lives better. Even though there has been a lot of work in treating cardiomyopathy, there are still some areas where more research is needed, especially for some forms or later stages of the disease. Researchers want to come up with new ways to treat illnesses, such as drug therapies, gene therapies, and stem cell-based treatments.

A lot of different types of people could benefit from research in this area, which could lead to better health results and fewer differences in health care. Some interesting new developments in regenerative medicine, like using stem cells and tissue engineering, may help fix heart tissue that has been harmed in people with cardiomyopathy. The goal of this research is to bring these new ideas into practical practice.

2. LITERATURE SURVEY

Heart disease is still one of the main reasons people die around the world, so we need good prediction models to help find and treat it early. Recently, methods called machine learning (ML) and deep learning (DL) have gotten a lot of attention because they might be able to correctly identify heart disease. This review of the literature looks at different studies and methods for using ML and DL models to identify heart disease.

Bakar et al. [1] give a thorough look at how ML and DL methods can be used to identify heart disease. The writers talk about how important it is to choose the right features, prepare the data, and choose the right model when making good prediction models. Their review shows the wide range of methods that were used, such as decision trees, support vector machines (SVMs), artificial neural networks (ANNs), and convolutional neural networks (CNNs).

A group of researchers, led by Dileep, [2], have come up with a way to automatically identify heart disease using a cluster-based bidirectional LSTM algorithm. By using the fact that LSTMs can work in both directions and grouping methods, the model does a good job of predicting heart disease, as shown by test results on large datasets.

Mishra et al. [3] talk about visual analysis methods for predicting cardiac arrest. They stress how important it is to raise knowledge and teach people about health. Their work shows that visual tools can help healthcare workers and the public understand complicated predictive information, which can lead to more preventative steps to avoid heart disease. Guo et al. [4] describe a motion-based analysis method for figuring out how long heart failure patients will live. The suggested model provides unique predictive ratings by looking at motion data. This lets healthcare workers make decisions that are best for each patient.

Nandy et al. [5] describe a swarm-artificial neural network (ANN)-based smart system for predicting heart disease. The model is more accurate at making predictions because it combines swarm intelligence with ANN design. This has been shown by a lot of testing on clinical datasets.

Pant et al. [6] suggest a new way to predict heart disease that uses picture segmentation with a CNN model. The CNN model accurately diagnoses and

predicts heart disease by separating cardiac photos into segments and pulling relevant features. This is a quick and painless way to solve the problem.

Saranya and Pravin [7] describe a way to choose features that uses both feature sensitivity and association analysis to make heart disease predictions more accurate. Their method makes it possible to find useful features while reducing the effects of redundant and unimportant factors. This makes prediction models more solid and easier to understand.

In conclusion, the literature review shows the wide range of ML and DL methods used to predict heart disease, from basic algorithms to advanced deep learning frameworks. These studies show how important it is to choose the right features, prepare the data, and make sure the model works best so that the predictions are correct and easy to understand. In addition, the use of motion-based observations, visual analysis, and new feature selection methods shows that research in predicting heart health is multidisciplinary, which has important effects for clinical practice and public health efforts.

3. METHODOLOGY

i) Proposed Work:

Through an interdisciplinary study method, the suggested system aims to improve how we understand and treat cardiomyopathy. The first step is to get an accurate collection with relevant traits and goal factors that show the presence of cardiomyopathy. Exploratory Data Analysis (EDA) shows how data is distributed, what values are missing, and how they are related. After that, data preparation gets the data ready for models by dealing with missing values, scaling features, and storing category variables. The system then uses Decision Tree and Random Forest algorithms that are already in place, setting their performance through hyperparameters. It also includes a K-Nearest Neighbors (KNN) algorithm that tries out different neighbor numbers to find the best one for performance. To rate and contrast the performance of the models, evaluation measures like accuracy, precision, recall, and F1-score are used. The goal of this complete system is to help researchers come up with better ways to diagnose and treat cardiomyopathy, which will eventually improve the lives of people who have it.

ii) System Architecture:

There are a few important steps in the system design for cardiomyopathy research. First, a cardiomyopathy dataset is gathered. Next, it is preprocessed to deal with missing values, scale features, and store category factors. After that, the data is split into training and testing sets to make evaluating the model easier. K-Nearest Neighbors (KNN), the Random Forest Classifier (RFC), and the Decision Tree Classifier (DTC) are the three classifiers that are used. Performance measures like accuracy, precision, recall, and F1-score are used to judge how well each model did after being taught on the training data. This method makes sure that a thorough review of the algorithms' ability to diagnose cardiomyopathy. The design makes it possible to create accurate and reliable detection tools for this difficult heart disease within a short time frame by carefully looking at the dataset, training models, and testing their performance.
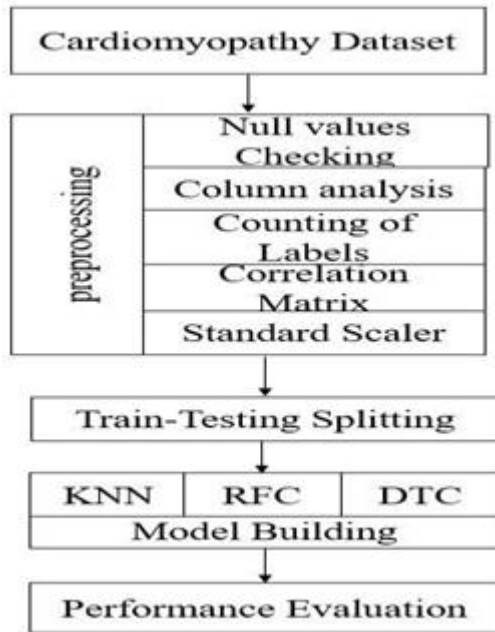
Fig 1 Proposed Architecture

iii) Dataset:

The collection comprises key heart health indicators. Since aging increases heart disease risk and males are more likely to have it, it starts with age and gender. Second, resting blood pressure, serum cholesterol, and fasting blood sugar are measured. All of these are crucial for heart health assessment. Additionally, clinical indications like chest pain (cp) and exercise-induced angina (exang) reveal patient complaints. The sleeping electrocardiogram (restecg), the greatest heart rate during exercise (thalach), and exercise-induced ST segment modifications (oldpeak, slope) are also crucial for diagnosis. Data on thalassemia (thal) and fluoroscopy-colored main blood vessels is also included. We learn more about the patient's heart. The aim variable (target) reveals if cardiac disease is present and is the major focus of forecasts. This strategy for gathering massive information helps us understand cardiomyopathy in various ways and develop effective diagnosis and treatment methods.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Fig 2 Dataset

iv) EDA:

Exploratory Data Analysis (EDA) was used to learn more about the cardiomyopathy dataset in this work. It was possible to see how the numbers were distributed by making histogram plots, which showed the range and frequency of values in each variable. The EDA gave a full picture of the dataset's numerical features by making a 15x15 figure and developing histograms for each number feature in the DataFrame "df."

A countplot was used to look at how the goal variable, which showed whether heart disease was present or not, was distributed. This image made it possible to see how balanced or skewed the information was toward different classes, which is very important for classification tasks.

In addition, an association image was made to show how the pairs of different traits are related. The heatmap showed possible connections between variables by finding the correlation matrix for each feature in the dataset and then picking the ones with the strongest correlations. The 'annot=True' option added numbers to the grid cells, which made it easier to see the size and direction of relationships. This made it easier to choose features and understand the model. Overall, this EDA method laid a solid basis for

further research and model building to deal with cardiomyopathy.

v) Data Processing:

Data pre-processing is an important first step in building a machine learning model, especially when working with real-world data that has noise, missing values, and different forms that machine learning algorithms can't understand. One-hot encoding is an important pre-processing method that turns category factors into a number format that machine learning models can use. This change, made easier by methods like pd.get_dummies(), makes binary columns for each group in categorical variables. This lets programs use categorical data effectively.

Standardization is another important step in pre-processing that can be done with tools like the Standard Scaler from scikit-learn. This step makes sure that the numbers are scaled so that the mean is 0 and the standard deviation is 1. Standardization has several benefits: it equalizes scaling across features, centers data around zero by taking away the mean, normalizes data by scaling with the standard deviation, and makes data easier to understand by putting features on the same scale, which makes it easier to compare how important different features are. By using one-hot encoding and standards to prepare the data, machine learning models can be more accurate and efficient, which makes their forecasting skills stronger and more reliable.

vi) Training & Testing:

Data pre-processing for machine learning involves separating the dataset into a training and test set to increase model performance and generalizability. We may test the model on a different data set after training it on the training set.

The machine learning model was taught using the training set. The input characteristics and output names are linked. This named data helps the model identify patterns, relationships, and links between input qualities and the objective variable.

The test set is a distinct dataset used to evaluate the model's performance and generalization. It gives the model new data to assess its future prediction abilities. By hiding this data during training, we make it seem that the model will encounter fresh data in real life.

Separating data into training and test sets tests the model's performance on new data. This shows its ability to foresee new scenarios better. This prevents overfitting and ensures the model can adjust to fresh data, improving its real-world reliability.

vii) Algorithms:

1. K-Nearest Neighbors (KNN):

Simple learning approach K-Nearest Neighbors (KNN) is used for classification and regression. A KNN-classified data point is the majority class. Regression uses the average of its k closest feature space neighbors. Set the hyperparameter k to inform the algorithm how many neighbors to look at. Projects may utilize KNN with Python tools like scikit-learn. Pre-processing comprises classifying characteristics size and encoding variables. Next, the KNN model is built up using factors like k neighbors. The model is taught from the training set. Finally, the model's test

set performance is assessed using F1-score, accuracy, precision, and recall.
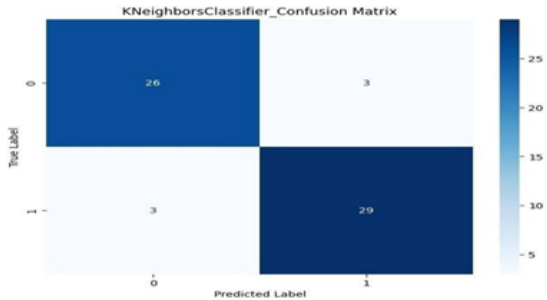
2. Random Forest (RF):

The Random Forest (RF) ensemble learning approach is used for regression and classification. It trains numerous decision trees and displays their mode (classification) or mean (regression) forecasts. Bootstrapping the training data and selecting a random set of features gives Random Forest its randomization. This reduces overfitting and improves generalization. Python's scikit-learn can utilize Random Forest. After cleaning the data, the Random Forest model is configured with hyperparameters like greatest depth and number of trees. Then, the training set instructs the model. After that, the model's test set performance is evaluated using the correct criteria.
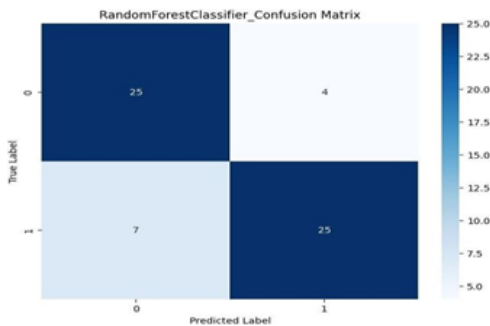
3. Decision Tree (DT):

Decision Tree (DT) directed learning may be used for classification and regression. Using input feature values, it recursively splits the feature space. A tree node represents each division. The decision tree chooses from the root node to a leaf node, where the outcome (class name or regression value) is. Decision Tree may be utilized with scikit-learn in Python. After preprocessing the data, the Decision Tree model is built up using hyperparameters like maximum depth and splitting rule. Then, the training set instructs the model. Finally, the model's test set performance is assessed using appropriate criteria.
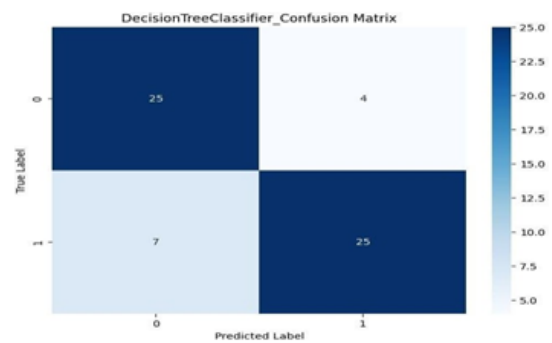
4. EXPERIMENTAL RESULTS

Correlation Matrix:

A correlation matrix is a way to look at two factors in a data set and figure out how they are related. Every cell in the matrix has a correlation value, with 1 meaning there is a strong relationship between the variables and 0 meaning there is no relationship at all and -1 meaning there is not a strong relationship.
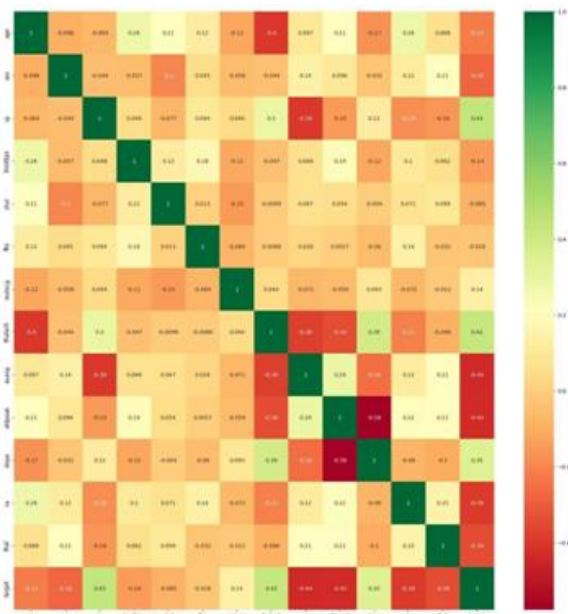
Fig 3 Correlation Matrix

Accuracy: How well a test can tell the difference between sick and healthy people is called its accuracy. To get an idea of how accurate a test is, we should figure out what percentage of cases are true positives and true negatives. In terms of math, this can be written as

Accuracy = TP + TN TP + TN + FP + FN.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Precision is the percentage of correctly classified cases or samples compared to those that were correctly classified as hits. So, here is the method to figure out the precision:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: In machine learning, recall is a parameter that shows how well a model can find all the important cases of a certain class. It shows how well a model captures cases of a certain class. It is calculated by dividing the number of correctly predicted positive observations by the total number of real positives.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: The F1 score is a way to rate the correctness of a machine learning model. It takes a model's accuracy and memory scores and adds them together. The accuracy measure counts how many times, across the whole collection, a model made a correct guess.

$$F1\ Score = \frac{2}{\left(\frac{1}{Precision} + \frac{1}{Recall}\right)}$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Class	Accuracy	Precision	Recall	F1–Score	Support
0	-	0.90	0.90	0.90	29
1	-	0.91	0.91	0.91	32
Macro Avg	-	0.90	0.90	0.90	61
Weighted Avg	0.90	0.90	0.90	0.90	61

Fig 4 KNN Classification report

Metric	Class 0	Class 1	Macro Avg	Weighted Avg
Precision	0.83	0.87	0.85	0.85
Recall	0.86	0.84	0.85	0.85
F1-Score	0.85	0.86	0.85	0.85
Support	29	32	61	61
Accuracy	85.25%	85.25%	85.25%	85.25%

Fig 5 Decision tree Classification report

Metric	Precision	Recall	F1-Score	Support
Class 0	0.83	0.86	0.85	29
Class 1	0.87	0.84	0.86	32
Macro Avg	0.85	0.85	0.85	61
Weighted Avg	0.85	0.85	0.85	61
Accuracy	0.85	0.85	0.85	

Fig 6 RFC Classification report

5. CONCLUSION

To sum up, this study's multifaceted analysis—which included detailed Exploratory Data Analysis (EDA), careful data preprocessing, and a full evaluation of the Decision Tree Classifier (DTC), the Random Forest Classifier (RFC), and the new k-Nearest Neighbors (KNN) Classifier—helped figure out how complicated the cardiomyopathy dataset was. This all-encompassing method taught us a lot about the dataset's natural properties, the complex ways each predictor worked, and their own strengths and weaknesses. The results of this study not only helped us learn more about how these machine learning models can be used to classify cardiomyopathy, but they also showed us possible areas for more research and improvement. These results could change the way heart health is diagnosed in the future by helping to make earlier diagnosis and treatment tools for cardiomyopathy more accurate and useful. This will eventually lead to better patient outcomes and better healthcare practices in this important area.

6. FUTURE SCOPE

The Future prospects for cardiomyopathy study and diagnosis are bright, and they include a wide range of new and improved areas. Magnetic resonance imaging (MRI) and ultrasound will continue to get better, giving us more accurate and detailed information about how the heart works. This will help in early diagnosis and personalized treatment plans. Heart muscle problems can be better diagnosed and treated based on a person's unique genetic background if we learn more about the genes that cause these problems.

Combining AI and machine learning techniques will make it more accurate to diagnose cardiomyopathy, predict risk, and suggest treatments by looking at large amounts of complicated data. New signs are being studied all the time so that cardiomyopathy can be found early and without surgery. This could lead to more effective treatments.

New discoveries in stem cells and tissue engineering could lead to regenerative therapies that can fix damaged heart muscle, which could reverse the effects of cardiomyopathy. Personalized treatment plans that take into account genetic, lifestyle, and external factors for each patient will lead to better therapy results and fewer side effects. As genetic tests and

technology get better, the laws and ethics that protect patient data privacy, genetic therapy, and informed consent will need to change too.

Cardiomyopathy studies and tests have a lot of room to grow in the future to make patient care, early diagnosis, and treatment results better. Researchers, healthcare workers, and people who are creating new technologies will need to keep working together to make these improvements happen and make sure they help patients all over the world.

REFERENCES

[1] Bakar, Wan Aezwani Wan Abu, et al. "A Review: Heart Disease Prediction in Machine Learning & Deep Learning." 2023 19th IEEE International Colloquium on Signal Processing & Its Applications (CSPA). IEEE, 2023.

[2] Dileep, P., et al. "An automatic heart disease prediction using cluster-based bidirectional LSTM (C-BiLSTM) algorithm." Neural Computing and Applications 35.10 (2023): 7253-7266.

[3] Mishra, Nilamadhab, et al. "Visual Analysis of Cardiac Arrest Prediction Using Machine Learning Algorithms: A Health Education Awareness Initiative." Handbook of Research on Instructional Technologies in Health Education and Allied Disciplines. IGI Global, 2023. 331-363.

[4] Guo, Saidi, et al. "Survival prediction of heart failure patients using motion-based analysis method." Computer Methods and Programs in Biomedicine 236 (2023): 107547.

[5] Nandy, Sudarshan, et al. "An intelligent heart disease prediction system based on swarm-artificial neural network." Neural Computing and Applications 35.20 (2023): 14723-14737.

[6] Pant, Aman, et al. "Heart disease prediction using image segmentation Through the CNN model." 2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2023.

[7]   Nandy, Sudarshan, et al. "An intelligent heart disease prediction system based on swarm-artificial neural network." Neural Computing and Applications 35.20 (2023): 14723-14737.

[8]   Saranya, G., and A. Pravin. "A novel feature selection approach with integrated feature sensitivity and feature correlation for improved prediction of heart disease." Journal of Ambient Intelligence and Humanized Computing 14.9 (2023): 12005- 12019.

[9]   Chandrasekhar, Nadikatla, and Samineni Peddakrishna. "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization." Processes 11.4 (2023): 1210.

[10]   Sherly, S. Irin, and G. Mathivanan. "An efficient honey badger based Faster region CNN for chronc heart Failure prediction." Biomedical Signal Processing and Control 79 (2023): 104165.

[11]   Rani, P., Kumar, R., Ahmed, N.M.O.S. et al. A decision support system for heart disease prediction based upon machine learning. J Reliable Intell Environ 7, 263–275 (2021).

[12]   M Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 1329-1333, doi: 10.1109/ICICT50816.2021.9358597.

[13]   S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 8154281554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[14]   Shah, D., Patel, S. & Bharti, S.K. Heart Disease Prediction using Machine Learning Techniques. SN COMPUT. SCI. 1, 345 (2020).

[15] C. Guo, J. Zhang, Y. Liu, Y. Xie, Z. Han and J. Yu, "Recursion Enhanced Random Forest with an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform," in IEEE Access, vol. 8, pp. 59247-59256, 2020, doi: 10.1109/ACCESS.2020.2981159.

[16]   Hager Ahmed, Eman M.G. Younis, Abdeltawab Hendawi, Abdelmgeid A. Ali, Heart disease identification from patients' social posts, machine learning solution on Spark, Future Generation Computer Systems, Volume 111, 2020, Pages 714-722, ISSN 0167-739X

[17]   Katarya, R., Meena, S.K. Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis. Health Technol. 11, 87–97 (2021).

[18]   Kannan, R., Vasanthi, V. (2019). Machine Learning Algorithms with ROC Curve for Predicting and Diagnosing the heart disease. In: Soft Computing and Medical Bioinformatics. SpringerBriefs in Applied Sciences and Technology (). Springer, Singapore.

[19]   Md Mamun Ali, Bikash Kumar Paul, Kawsar Ahmed, Francis M Bui, Julian M.W. Quinn, Mohammad Ali Moni, Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison, Computers in Biology and Medicine, Volume 136, 2021, 104672, ISSN 0010-4825.

[20]   Ibomoiye Domor Mienye, Yanxia Sun, Zenghui Wang, An improved ensemble learning approach for the prediction of heart disease risk, Informatics in Medicine Unlocked, Volume 20, 2020, 100402, ISSN 2352-9148.

[21]   Aniruddha Dutta, Tamal Batabyal, Meheli Basu, Scott T. Acton, An efficient convolutional neural network for coronary heart disease prediction, Expert Systems with Applications, Volume 159, 2020, 113408, ISSN 0957-4174.

[22]   C. Beulah Christalin Latha, S. Carolin Jeeva, Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques, Informatics in Medicine Unlocked, Volume 16, 2019, 100203, ISSN 2352-9148.

[23]   A. Ishaq. "Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques," in IEEE Access, vol. 9, pp. 3970739716, 2021, doi: 10.1109/ACCESS.2021.3064084.

[24]   Shahrokh Asadi, SeyedEhsan Roshan, Michael W. Kattan, Random Forest swarm optimization-based for heart diseases diagnosis, Journal of Biomedical Informatics, Volume 115, 2021, 103690, ISSN 1532-0464.