Prediction of Cardiovascular Disease using Machine Learning Algorithms with Relief and Lasso Feature Selection Techniques

Ms. Lakshmi Devi Kadali, M Tech student, CSE:BVCE, lakshmikadali051197@gmail.com
Prof. V. S. Ramakrishna, Associate Professor, CSE:BVCE, vsramakrishna.bvce@bvcgroup.in
Dr. Chandra Mouli VSA, Prof CSE:BVCE, mouliac@yahoo.co.in
Dr. Rajasekhar, Prof ECE:BVCE, raja.bubbly@gmail.com
Gunamani Jena, Prof CSE:BVCE, drgjena@gmail.com

Article Info
Page Number: 5356 - 5372
Publication Issue:
Vol 71 No. 4 (2022)

Abstract

Cardiovascular diseases (CVD) are one of the most common types of serious illnesses. Early diagnosis can help stop or lessen the effects of CVDs, which could lower death rates. Using machine learning models to find risk factors is a promising idea. We'd like to suggest a model that uses different methods to predict heart disease more accurately. For our proposed model to work, we used efficient methods for Data Collection, Data Pre-processing, and Data Transformation to get accurate data for training the model. We used a group of datasets (Cleveland, Long Beach VA, Switzerland, Hungarian and Stat log). The Relief and Least Absolute Shrinkage and Selection Operator (LASSO) techniques are used to choose the right features. By combining traditional classifiers with bagging and boosting methods, which are used in the training process, new hybrid classifiers like Decision Tree Bagging Method (DTBM), Random Forest Bagging Method (RFBM), K-Nearest Neighbors Bagging Method (KNNBM), AdaBoost Boosting Method (ABBM), and Gradient Boosting Boosting Method (GBBM) are made. We also used some machine learning algorithms to figure out the Accuracy (ACC), Sensitivity (SEN), Error Rate, Precision (PRE), and F1 Score (F1) of our model, as well as the Negative Predictive Value (NPR), False Positive Rate (FPR), and False Negative Rate (FNR). We used Convolution 2D Neural Networks because other algorithms didn't give the best results and this algorithm is a deep learning and advanced version of all existing

Article History
Article Received*: 25 March 2022*
Revised*: 30 April 2022*
Accepted*: 15 June 2022*
Publication*: 19 August 2022*

Heart disease, machine learning, cardiovascular disease (CVD), Relief feature selection, LASSO feature selection, decision tree, random forest, K-nearest neighbours, AdaBoost, and gradient boosting are all words that describe these things.

I. INTRODUCTION:

Cardiovascular disease has long been thought to be the most dangerous and deadly disease that humans can get. The high death rate from cardiovascular diseases is putting a lot of stress and risk on healthcare systems all over the world. Men are more likely to get cardiovascular diseases than

women, especially when they are middle-aged or older [1, 2]. However, children can also have the same health problems. WHO data show that heart disease is the cause of one-third of all deaths around the world. About 17.9 million people die every year from CVDs around the world, and they are more common in Asia [4, 5]. The European Cardiology Society (ESC) said that heart disease has been found in 26 million adults around the world, and 3.6 million new cases are found each year. About half of the people who are diagnosed with heart disease die within 1–2 years, and about 3% of the total health care budget goes toward treating heart disease [6]. To predict heart disease, you need to do more than one test. If the medical staff doesn't know enough, they might make wrong predictions [7]. It can be hard to find a problem early [8]. Heart disease is hard to treat with surgery, especially in developing countries where there isn't enough trained medical staff, testing equipment, and other resources to properly diagnose and care for people with heart problems [9]. Patients would be safer and heart attacks wouldn't be as bad if the risk of cardiac failure was accurately assessed [10]. Machine learning algorithms can be used to find diseases if they are taught with the right data [11]. There are public datasets on heart disease that can be used to compare different prediction models. With the help of machine learning and artificial intelligence, researchers can use the large databases they have to make the best prediction model possible. Recent studies on heart problems in adults and children have shown how important it is to cut down on deaths caused by cardiovascular diseases (CVDs). Since the clinical datasets that are already out there are inconsistent and duplicate, preprocessing is a very important step [12]. It is important to choose the important features that can be used as risk factors in prediction models. To make accurate prediction models, it is important to choose the right combination of the features and the right machine learning algorithms [13]. It is important to look at the effects of risk factors that meet three criteria, such as being common in most populations, having a big effect on heart disease on their own, and being able to be controlled or treated to lower the risks [14]. When predicting CVD, different researchers have used different risk factors or characteristics. Age, sex, chest pain (cp), fasting blood sugar (FBS)—high FBS is linked to Diabetes [72], resting electrocardiographic results (Restecg), exercise-induced angina (exang), ST depression induced by exercise relative to rest (oldpeak), slope, number of major vessels coloured by fluoroscopy (ca), heart status (thal), maximum heart rate achieved (thalach), poor diet, family history, and maximum heart rate achieved (thalach) are some of the factors Recent studies show that the prediction needs at least 14 things to be accurate and reliable [20]. Researchers have a hard time putting these factors together with the right machine learning methods to make a good prediction of heart disease [21]. Machine learning algorithms work best when they are trained on the right sets of data [22–25]. Since the algorithms depend on the training and test data being the same, using feature selection techniques like data mining, Relief selection, and LASSO can help prepare the data so that a more accurate prediction can be made. Once the important features have been chosen, classifiers and hybrid models can be used to estimate how likely it is that a disease will happen. Different methods have been used by researchers to make classifiers and hybrid models [12, 20]. There are still a number of things that might make it hard to accurately predict heart disease, such as a lack of in-depth analysis, limited medical datasets, feature selection, and ML algorithm applications.

1.1 Cardiovascular diseases:

Heart and blood vessel diseases (CVDs) kill an estimated 17.9 million people around the world every year. CVDs are a group of diseases that affect the heart and blood vessels. They include coronary heart disease, cerebrovascular disease, rheumatic heart disease, and other conditions. More than four out of five deaths from CVD are caused by heart attacks and strokes, and a third of these deaths happen too soon in people under the age of 70.

The most important things that people do that put them at risk for heart disease and stroke are eating poorly, not being active, smoking, and drinking too much. Behaviors that put people at risk can cause health problems like high blood pressure, high blood sugar, high blood lipids, and being overweight or obese. These "intermediate risk factors" can be measured in primary care centres and show that a person is more likely to have a heart attack, stroke, heart failure, or other problem.

Heart disease is less likely to happen if you stop smoking, eat less salt, eat more fruits and vegetables, work out regularly, and don't drink too much. Also, drinking too much alcohol can hurt your heart. Health policies that make it easy and affordable for people to make healthy choices are important for getting people to start and keep up healthy habits.

As the number of people grows, so do the chances of getting sick. There are many diseases in the world, and one of the biggest problems hospitals face today is that they don't have the technology to tell when a patient is sick. CVD, which stands for Cardiovascular Disease, is one of these. It means any kind of heart, blood vessel, or blood vessel disease. WHO says that more people die from CVDs than from any other cause. It hurts more the countries with low and middle incomes. When they're sick, it's hard for people who live alone to call the hospital. So, we've made a model that can tell when a patient is sick and tell the hospital about it. At the moment, the system only finds people with heart disease and tells the hospital. We chose heart disease identification because it is one of the most dangerous diseases and a high number of people die from it every year.

1.2 Relief:

Kira and Rendell made the first Relief algorithm, which was based on the idea of "learning by doing." Relief calculates a proxy statistic for each feature that can be used to estimate the "quality" or "relevance" of the feature to the target concept. This is a method of individual evaluation filtering feature selection (i.e. predicting endpoint value). Relief has been described as both non-myopic and non-parametric, which means it doesn't make any assumptions about the size or distribution of the population or the sample. People have said that the algorithm works well because it doesn't explicitly look at feature subsets and doesn't try to find an optimal minimum feature subset size.

LASSO stands for Least Absolute Shrinkage and Selection Operator.

LASSO is a way to do regression analysis that uses attribute selection and regularisation to make the final statistical model easier to predict and understand.

1.3 Motivation:

Heart disease kills a lot of people every year all over the world. To lower this death rate, we need to find ways to diagnose heart disease symptoms quickly and accurately.

1.4 The World Health Organization says that heart disease is the cause of one-third of all deaths around the world. About 17.9 million people die every year from CVDs around the world, and they are more common in Asia [4, 5]. The European Cardiology Society (ESC) said that heart disease has been found in 26 million adults around the world, and 3.6 million new cases are found each year. About half of the people who are diagnosed with heart disease die within 1–2 years, and about 3% of the total health care budget goes toward treating heart disease [6]. To predict heart disease, you need to do more than one test. If the medical staff doesn't know enough, they might make wrong predictions [7]. It can be hard to find a problem early [8]. Heart disease is hard to treat surgically, especially in developing countries where there isn't enough trained medical staff, testing equipment, and other tools needed to find out what's wrong with a patient and take care of them properly.

1.5 The main goal of this article is to fill in some of these gaps in research so that a better model for predicting CVD can be made. In this study, Relief and LASSO are used to pick out the most important features from medical references based on their rank values. This also helps with the machine learning problems of overfitting and underfitting.

1.6 Work Scope:

Features related to the most important risk factors were ranked, and traditional biostatistics tests and the machine learning algorithms were put side by side and compared. All of the existing algorithms (DTBM, RFBM, KNNBM, ABBM, and GBBM) focus on tuning machine learning algorithms with different input parameters. LASOO and Relief features selection algorithms have not been used in any existing work.

1.7 Order of the Report:

Five chapters make up the rest of the report. After this first chapter, chapter 2 talks about a survey of the system that is already in place. This gives an overview of the research that has been done so far to find ways to quickly and accurately diagnose heart-related symptoms. In this paper, the author talks about hybrid machine learning with the help of the Relief and Lasso features selection algorithms. Lasso (Least Absolute Shrinkage and Selection Operator) and Relief help get important features from the training dataset. These features are then trained with hybrid machine learning algorithms such as Decision Tree Bagging Method (DTBM), Random Forest Bagging Method (RFBM),K-Nearest Neighbors Bagging Method (KNNBM), AdaBoost Boosting Method (ABBM), and GradientBoosting Boosting Method (GBBM). The performance of each algorithm is judged based on its Accuracy, Sensitivity, Precision, TPR, and FPR.

In Chapter 3, the proposed system is explained. This begins with an explanation of the dataset and the models used in the report. Then it talks about how the proposed system is put together. Describes the process and algorithms that were used, as well as the details of the software that was used for the research. It also talks about the criteria that were used to evaluate this study.

Chapter 4 tells about the experiment and how it turned out. Smart healthcare systems can use this model to find the right way to diagnose diseases.

In Chapter 5, the results of all the models in this research paper are summed up, and suggestions are made about when to use each model. It shows how work should be done in the future.

2.2 Research Contribution: In this paper, Decision Tree, Random Forest, KNN, AdaBoost, and Gradient Boosting are made by combining traditional classifiers with bagging and boosting methods, which are used in the training process. The performance of each algorithm is judged based on its Accuracy, Sensitivity, Precision, TPR, and FPR.

III. System Planned

In the work we are proposing, we are using a combination of Lasso and Relief features selection with hybrid machine learning algorithms. Selected features will have a strong relationship with each other, so training with hybrid algorithms helps improve accuracy.

3.1 Algorithms/ techniques:

In this study, we use the Decision Tree, Random Forest, KNN, AdaBoost, Gradient Boosting, and Convolution 2D Neural Networks algorithms.

3.1.1 Algorithm for a random forest:

This model is based on three random ideas: picking training data at random when making trees, picking some subsets of features when splitting nodes, and only looking at a subset of all features when splitting each node in each simple decision tree. In a random forest, each tree learns from a random sample of the data points when it is being trained. A random forest model is made up of a large number of decision trees. The model basically takes the average of what the trees predict, which is why it is called a forest. Also, the algorithm includes three random ideas: picking training data at random when making trees, picking some subsets of variables at random when splitting nodes, and deciding that only a subset of all variables should be used to split every node in each basic decision tree. During the training of a random forest, each basic tree learns from a random sample of the dataset.

3.1.2 Decision Tree Algorithm: This algorithm builds a training model by putting all similar records on the same branch of a tree. This process continues until all records are in the tree. Classification train model will be used to describe the whole tree.

3.1.3 Gradient Boosting Algorithm: Gradient boosting classifiers are a group of machine learning algorithms that combine many weak learning models into one strong predictive model. When gradient boosting is done, decision trees are often used. Gradient boosting models are becoming more popular because they can classify complex datasets well. Recently, they have been used to win a lot of Kaggle data science competitions.

3.1.4 KNN (K-Nearest Neighbor) algorithm: KNN is usually thought to have two properties: lazy learning and a non-parametric algorithm. This is because KNN doesn't make any assumptions about how the underlying data is spread out. To find targets, the method has a few steps: Step 1: Start

Step 2: Separating the data into training data and test data

Step 3: Choose a value for K and figure out which distance function to use.

Step 4: Choose a test data sample.

Step 5: stop 3.1.5 Algorithm for AdaBoost

AdaBoost, which is also known as "Adaptive Boosting," is a Machine Learning method that is used as an Ensemble Method. Most of the time, AdaBoost is used with decision trees with only one level, which means that there is only one split. People also call these trees "Decision Stumps."

3.2Architecture/Framework:

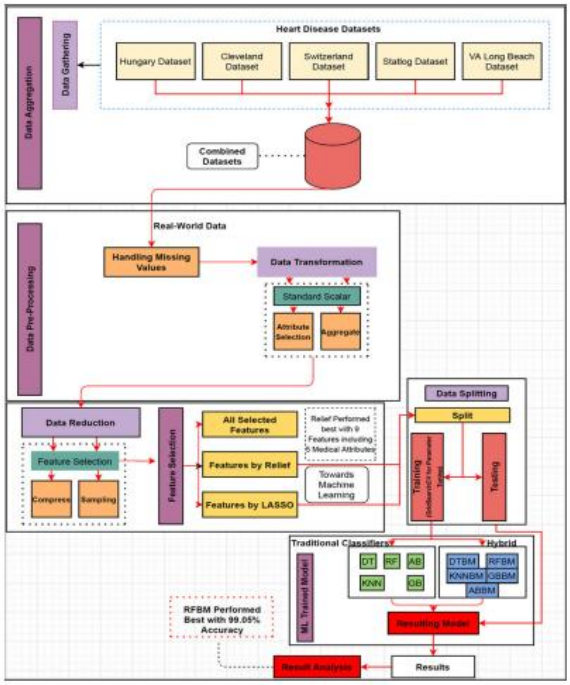

Fig.1 Architecture/Framework

3.3Algorithm and Process Design:

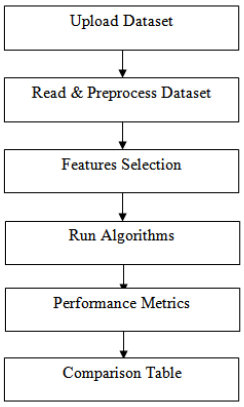

Fig 2. Algorithm and process design

1) Upload Dataset: We will upload the heart statlogclevelandhungary final dataset using this module.

2) Read and prepare dataset: We are defining the MIN-MAX scaler object, reading dataset, replacing missing values with 0, and then printing dataset values.

3) Selecting Features: Use the LASSO algorithm to pick out the most important parts of the dataset. Dataset has a total of 11 features. After applying LASSO, we chose 6 features, and those 6 features are:

4) Run Decision Tree, Random Forest, and KNN, and then combine Boosting classifier with AdaBoost and Gradient Boosting, Convolution 2D Neural Networks. The results of the experiments show that all of the algorithms with Relief features selection, CNN, and it got high accuracy and CNN got less error.

5) Performance Metrics: In this module, we will plot the accuracy of both algorithms.

6) Comparison Table: With this module, we'll show how all algorithms compare to each other.

4 How It Was Done and What Happened

4.1 Getting information

Researchers use the https://www.kaggle.com/datasets/sid321axn/heart-statlog-cleveland-hungary-final dataset the most.

Heart disease, also called cardiovascular diseases (CVDs), is the leading cause of death around the world. About 17.9 million people die of heart disease every year, which is about 32% of all deaths. CVDs are a group of heart and blood vessel diseases that include coronary heart disease, cerebrovascular disease, rheumatic heart disease, and other conditions. Heart attacks and strokes

cause four out of every five deaths from CVD, and one-third of these deaths happen too soon in people under the age of 70.

We put together this dataset by putting together different datasets that were already out there but had never been put together before. We've put them all together based on 11 things they have in common, making it the largest research dataset on heart disease.

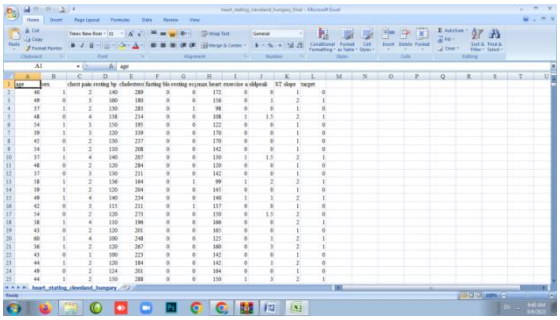

Fig 3: Data Collection

4.2 Evaluation Metrics:

F1-Score, Accuracy, and Receivers Operating Characteristics-Area We use the Area Under the Curve (ROC-AUC) metric to measure how well our models work. F1-score, Accuracy, Precision, and Recall must be evaluated by: • FPR = False Positive Rate

• TPR stands for "True Positive Rate"

• Correctness • Accuracy • Preciseness • Memory • F1-score

For this, values are calculated based on: • True positive (TP) = number of events for which the correct answer was given.

False negative (FN): The number of events that were wrongly predicted and didn't happen. False positive (FP): The number of events that were wrongly predicted.

• True negative (TN): The number of events that could have happened but didn't.

False Positive Rate (FPR): This is a way to measure how accurate machine learning is. Its formula is: FPR = FP / (FP + TN)

Rate of True Positives (TPR): It is the same as recall, so its definition is TPR=FP/(FP+TN).

Accuracy is the most important way to measure performance, and it's easy to do with a ratio of the number of correct predictions to the total number of observations.

Accuracy=(TN+TP)/(TP+FP+TN+FN)

Remember that it is the number of positive observations that can be predicted correctly out of all the observations in the original data.

Recall = (TP+FN)/(TP+TP)

Precision: It is used to figure out which values are correct. This means to take the total number of software's that were predicted as positive and figure out how many of them were right. Precision = TP / (TP + FP) says what it means.

F1-score: The F-score is a way to combine the model's precision and recall. The mean of the model's precision and recall is used to calculate the F-score. It's also known as a "F-score." It is written as F1 Score = 2 (Precision Recall/Precision + Recall).

The area under the ROC-AUC curve from prediction scores is used to figure out the ROC-AUC metric, which is another powerful way to solve classification problems.

4.3 Outcome:

We used JUPYTER Notebook to finish this project, and the results are shown below.

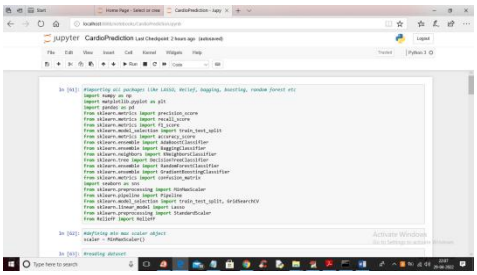

Fig 4: Importing all required packages

In above figure we have wrote comments starting with # symbol and in above figure we are importing all required packages

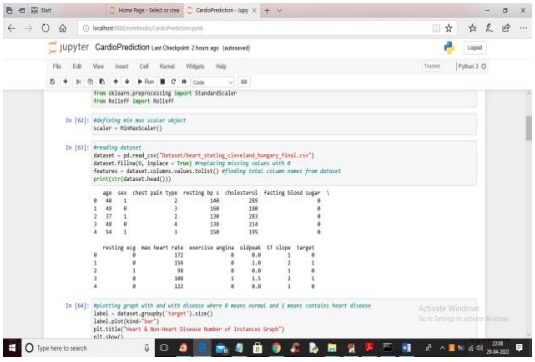

Fig 5: MIN-MAX scaler

In above figure we are defining MIN-MAX scaler object and then reading dataset and then replacing missing values with 0 and then printing dataset values

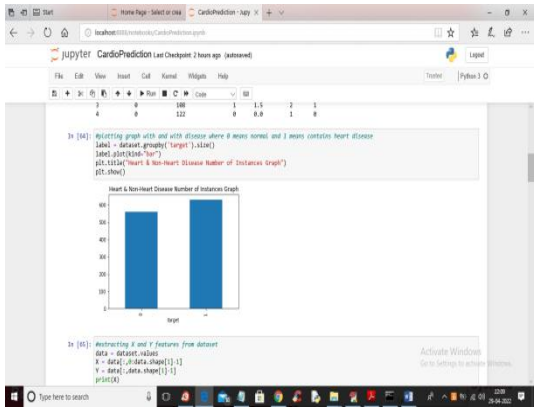

Fig 6:Normal and Heart disease no. of records found dataset

In above figure we are plotting normal and heart disease total number of records found in dataset and in above graph x-axis represents 0 (normal) and 1 (heart disease) and y-axis represents number of records found in 0 and 1 class.

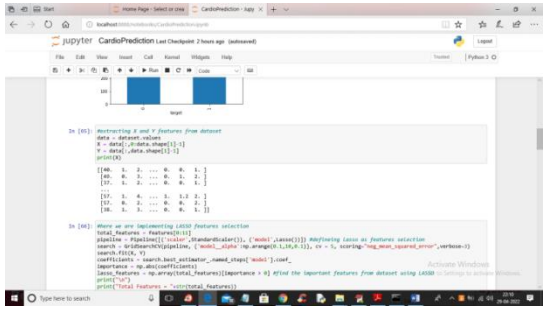

Fig 7: Extracting X&Y features from dataset

In above figure we are extracting X and Y features from dataset and then printing those values

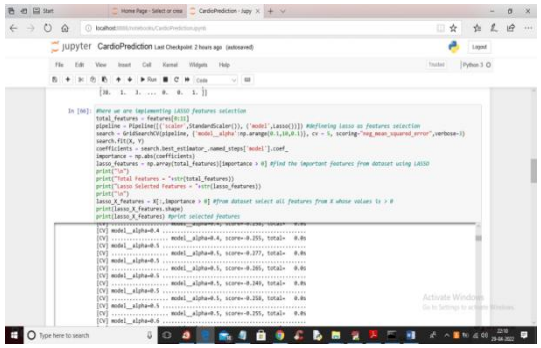

Fig 8: LASSO features selection

In above figure we wrote comments on how to use LASSO algorithm to select important features from dataset and after running that block will get below selected features

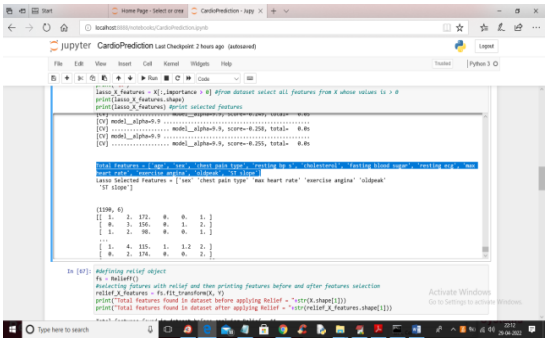

Fig 9: After applying LASSO

In above figure in blue colour text you can see dataset contains total features as 11 and after applying LASSO we got 6 selected features and those selected features you can see after blue line.

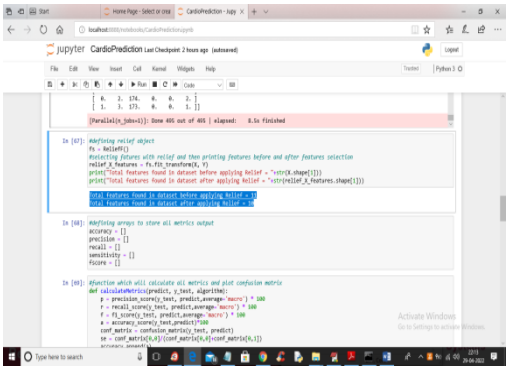

Fig 10: defining array for accuracy and other metrics

In above figure in blue colour text you can see features selected with Relief where total features are 11 and Relief selected 10 features from it and in next block we define arrays for accuracy and other metrics

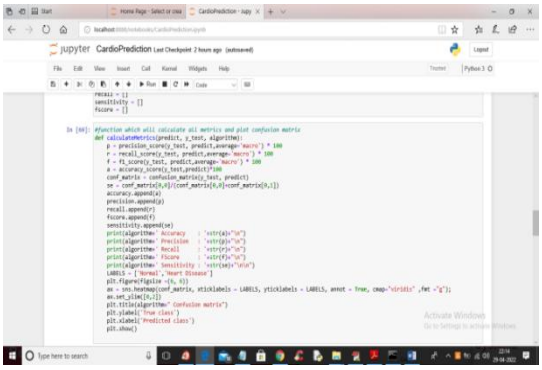

Fig 11:function to calculate accuracy and other metrics.

In above figure we wrote function to calculate accuracy and other metrics for each algorithm

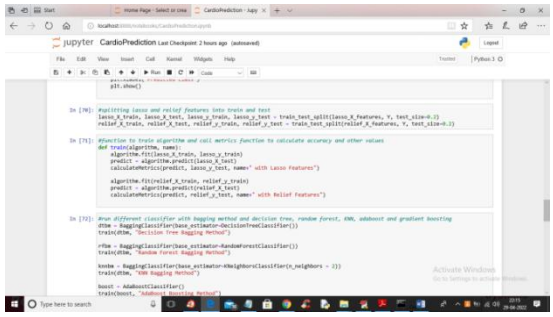

Fig 12:splitting both LASSO & RELIEF

In above figure in first block we are splitting both LASSO and RELIEF features into train and test and then defining 'train' function to train all algorithms and in last block we create object of each bagging and boosting classifier and then call train function to perform training on train and testing on test data and after running above blocks will get below output for each algorithm

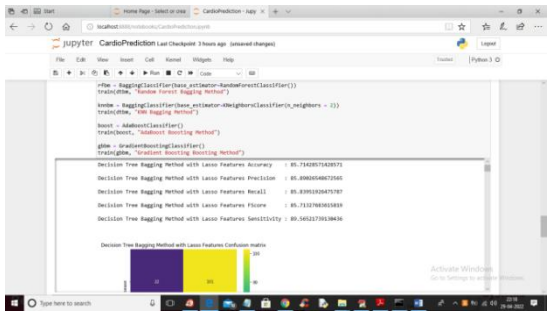

Fig 13: bagging decision tree performance with LASSO

In above figure we can see bagging decision tree performance with LASSO features and in below figure we can see decision tree with RELIEF

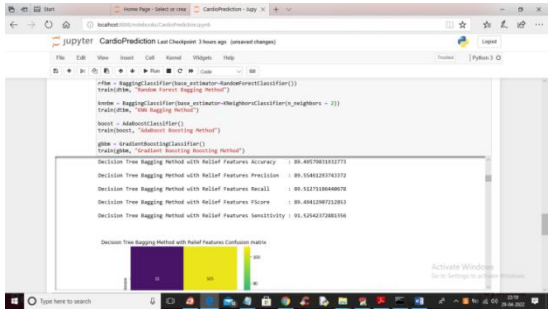

Fig 14:decision tree with RELIEF

In below 2 figures we can see bagging Random Forest with LASSO and RELIEF

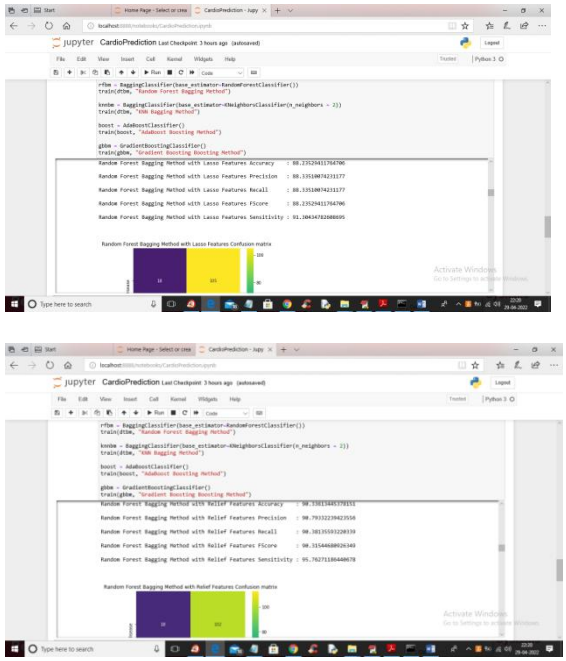

Fig 15: Random Forest with LASSO & RELIEF

In below 2 figures you can bagging KNN output with LASSO and RELIEF

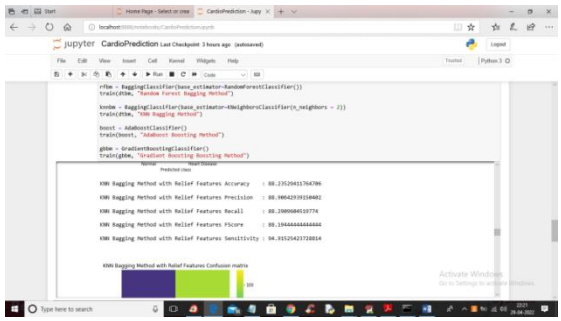

Fig 16:KNN output with LASSO and RELIEF

In below 2 figures you can see ADABOOST boosting output for LASSO and RELIEF

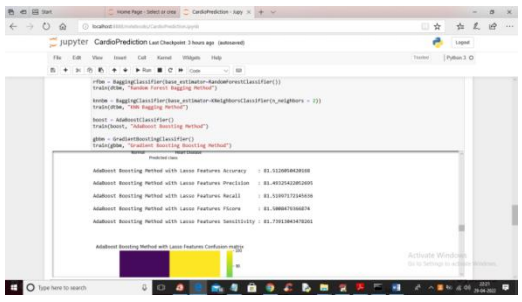

Fig 17:ADABOOST boosting output for LASSO and RELIEF

In below 2 figures you can see gradient boosting output with LASSO and RELIEF

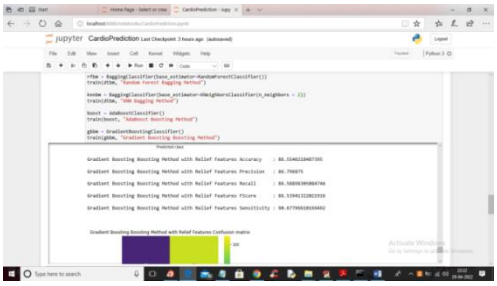

Fig 18:gradient boosting output with LASSO and RELIEF

From above output we can confirm that RELIEF is giving better result compare to LASSO and below is the performance of all algorithms

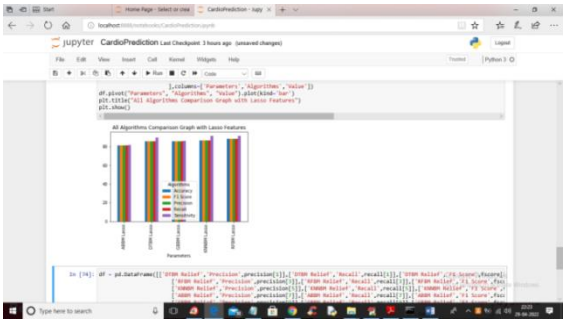

Fig 19:all algorithms performance with LASSO features

In above graph we can see all algorithms performance with LASSO features and in above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics values where each metric represents different colour bar and below is the performance of all algorithms with RELIEF features

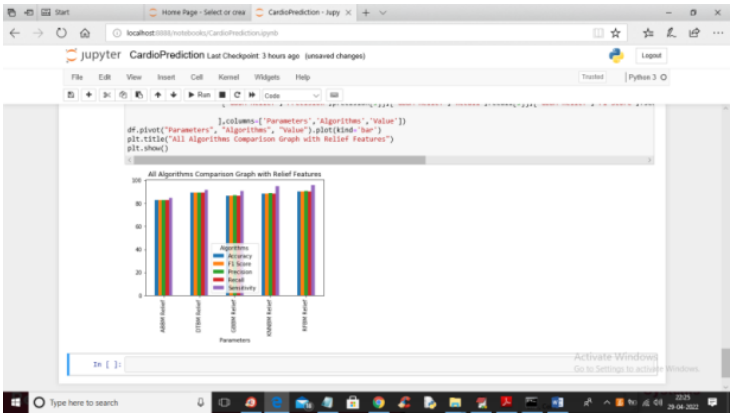

Fig 20:all algorithms with RELIEF features

Above graph is for RELIEF features

Extension outcomes:

Cardiovascular Extension

In this project as extension we have used Convolution 2D Neural Networks as other algorithms are not giving best result and this algorithm is a deep learning and advance version of all existing Machine Learning algorithms. In below figure we are showing accuracy of all algorithms

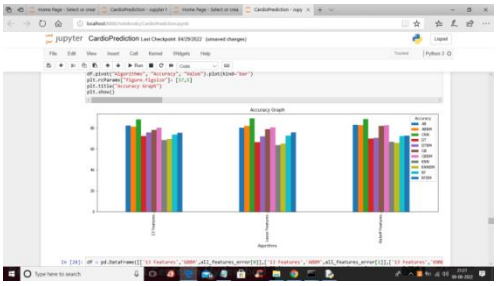

Fig 21:accuracy of all algorithms

In above figure x-axis represents Features Type and y-axis represents accuracy and in all colours bar Green colour refers to extension CNN and it got high accuracy and similarly we got less error for CNN which you can see in below figure

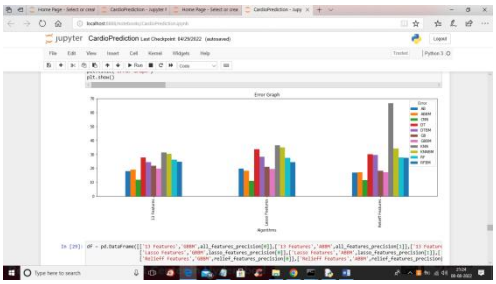

Fig 22:less error compare to existing and propose algorithms.

In the error graph above, the green bar represents CNN, which made less mistakes than both existing algorithms and proposed algorithms. In the same way, you can look at all graphs of metrics in your Notebook or HTML output file.

CONCLUSION

Identifying the risk of heart disease with a fair amount of accuracy could have a big impact on the long-term death rate of all people, no matter their culture or social status. Getting a diagnosis early is a key part of reaching this goal. With the help of machine learning, heart disease has already been tried to be predicted in a few studies. This study goes in a similar direction, but it uses a new and improved method and a bigger set of data to train the model. This research shows that the Relief feature selection algorithm can provide a set of closely related features that can then be used with several machine learning algorithms. Other algorithms weren't giving us the best results, so we used

Convolution 2D Neural Networks as an extension. This algorithm is a deep learning and advanced version of all the other Machine Learning algorithms.

Bibliography

[1] C. Trevisan, G. Sergi, S. J. B. Maggi, and H. Dynamics, ‘‘Gender differences in brain-heart connection,’’ in Brain and Heart Dynamics. Cham, Switzerland: Springer, 2020, p. 937.

[2] M. S. Oh and M. H. Jeong, ‘‘Sex differences in cardiovascular disease risk factors among Korean adults,’’ Korean J. Med., vol. 95, no. 4, pp. 266–275, Aug. 2020.

[3] D. C. Yadav and S. Pal, ‘‘Prediction of heart disease using feature selection and random forest ensemble method,’’ Int. J. Pharmaceutical Res., vol. 12, no. 4, 2020.

[4] World Health Organization and J. Dostupno, ‘‘Cardiovascular diseases: Key facts,’’ vol. 13, no. 2016, p. 6, 2016. [Online]. Available: https:// www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases- (cvds)

[5] K. Uyar and A. Ilhan, ‘‘Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks,’’ Procedia Comput. Sci., vol. 120, pp. 588–593, Jan. 2017.

[6] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, ‘‘A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms,’’ Mobile Inf. Syst., vol. 2018, pp. 1–21, Dec. 2018.

[7] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia, and J. Gutierrez, ‘‘A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease,’’ in Proc. IEEE Symp. Comput. Commun. (ISCC), Jul. 2017, pp. 204–207.

[8] J. Mourao-Miranda, A. L. W. Bokde, C. Born, H. Hampel, and M. Stetter, ‘‘Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data,’’ NeuroImage, vol. 28, no. 4, pp. 980–995, Dec. 2005.

[9] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, ‘‘Innovative artificial neural networks-based decision support system for heart diseases diagnosis,’’ J. Intell. Learn. Syst. Appl., vol. 5, no. 3, pp. 176–183, 2013.

[10] Q. K. Al-Shayea, ‘‘Artificial neural networks in medical diagnosis,’’ Int. J. Comput. Sci., vol. 8, no. 2, pp. 150–154, 2011.

[11] F. M. J. M. Shamrat, M. A. Raihan, A. K. M. S. Rahman, I. Mahmud, and R. Akter, ‘‘An analysis on breast disease prediction using machine learning approaches,’’ Int. J. Sci. Technol. Res., vol. 9, no. 2, pp. 2450–2455, Feb. 2020.

[12] M. S. Amin, Y. K. Chiam, and K. D. Varathan, ‘‘Identification of significant features and data mining techniques in predicting heart disease,’’ Telematics Informat., vol. 36, pp. 82–93, Mar. 2019.

[13] N. Kausar, S. Palaniappan, B. B. Samir, A. Abdullah, and N. Dey, ‘‘Systematic analysis of applied data mining based optimization algorithms in clinical attribute extraction and classification for diagnosis of cardiac patients,’’ in Applications of Intelligent Optimization in Biology and Medicine. Cham, Switzerland: Springer, 2016, pp. 217–231.

[14] J. Mackay and G. A. Mensah, ‘‘The atlas of heart disease and stroke,’’ World Health Org., Geneva, Switzerland, Tech. Rep., 2004.

[15] M. Ashraf, S. M. Ahmad, N. A. Ganai, R. A. Shah, M. Zaman, S. A. Khan, and A. A. Shah, Prediction of Cardiovascular Disease Through Cutting-Edge Deep Learning Technologies: An Empirical Study Based on TENSORFLOW, PYTORCH and KERAS. Singapore: Springer, 2021, pp. 239–255.

[16] F. Andreotti, F. S. Heldt, B. Abu-Jamous, M. Li, A. Javer, O. Carr, S. Jovanovic, N. Lipunova, B. Irving, R. T. Khan, R. Dürichen, ‘‘Prediction of the onset of cardiovascular diseases from electronic health records using multi-task gated recurrent units,’’ 2020, arXiv:2007.08491. [Online]. Available: https://arxiv.org/abs/2007.08491

[17] W. Wiharto, H. Kusnanto, and H. Herianto, ‘‘Hybrid system of tiered multivariate analysis and artificial neural network for coronary heart disease diagnosis,’’ Int. J. Electr. Comput. Eng., vol. 7, no. 2, p. 1023, Apr. 2017.

[18] A. K. Paul, P. C. Shill, M. R. I. Rabin, and M. A. H. Akhand, ‘‘Genetic algorithm based fuzzy decision support system for the diagnosis of heart disease,’’ in Proc. 5th Int. Conf. Informat., Electron. Vis. (ICIEV), May 2016, pp. 145–150.

[19] X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang, and Q. Wang, ‘‘A hybrid classification system for heart disease diagnosis based on the RFRS method,’’ Comput. Math. Med., vol. 2017, pp. 1–11, Jan. 2017.

[20] D. Singh and J. S. Samagh, ‘‘A comprehensive review of heart disease prediction using machine learning,’’ J. Crit. Rev., vol. 7, no. 12, p. 2020