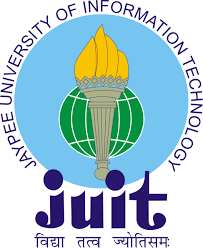
Project Report

on

Autonomous Image Caption Generator using Neural Network

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



Submitted by-

181209     181237

Apurva Dubey Sudhanshu Shridhar

**DECLARATION**

I hereby declare that this project has been done by me under the supervision of **Dr. Ruchi Verma**, **Assistant Professor**, Department of CSE and IT, Jaypee University of Information Technology. I also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

**(Dr. Ruchi Verma)**

**Assistant Professor**

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology

Submitted by:

**(Apurva Dubey) (Sudhanshu Shridhar)**

**(181209) (181237)**

Computer Science & Engineering Department

Jaypee University of Information Technology

**CERTIFICATE**

This is to certify that the work which is being presented in the project report titled “**Risk Stratification Prior To PCI**” is in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science And Engineering and submitted to the Department of Computer Science And Engineering, Jaypee University of Information Technology, Waknaghat is an authentic record of work carried out by “**Apurva Dubey (181209), Sudhanshu Shridhar (181237)**” during the period from August 2021 to December 2021 under the supervision of **Dr. Ruchi Verma**, Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat.

**Apurva Dubey Sudhanshu Shridhar**

**(181209) (181237)**

The above statement made is correct to the best of my knowledge.

**(Dr. Ruchi Verma)**

**Assistant Professor**

Computer Science & Engineering and Information Technology

Jaypee University of Information Technology, Waknaghat

**ACKNOWLEDGEMENT**

Firstly, I express my heartiest thanks and gratefulness to Almighty God for his divine blessing that makes it possible to complete the project work successfully.

I am really grateful and wish my profound indebtedness to Supervisor **Dr. Ruchi Verma**, **Assistant Professor**, Department of CSE Jaypee University of Information Technology, Wakhnaghat. Deep Knowledge & keen interest of my supervisor in the field of **“Artificial Intelligence”** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

I would like to express my heartiest gratitude to **Dr. Ruchi Verma**, Department of CSE, for his kind help to finish my project.

I would also generously welcome each one of those individuals who have helped me straightforwardly or in a roundabout way in making this project a win. In this unique situation, I might want to thank the various staff individuals, both educating and non-instructing, which have developed their convenient help and facilitated my undertaking.

Finally, I must acknowledge with due respect the constant support and patience of my parents.

**Apurva Dubey**

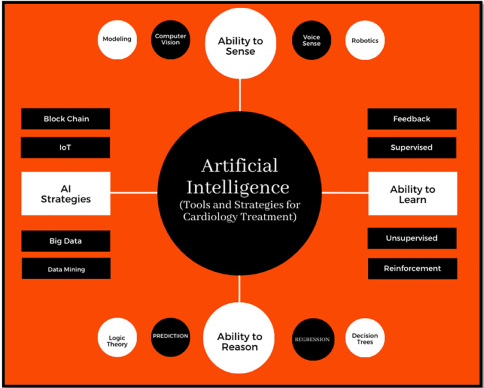
**Sudhanshu Shridhar**

TABLE OF CONTENT

1. Abstract 6
2. Introduction 7
3. Literature Survey 8
4. Model Description 23
5. Data Description 29
6. Data Preprocessing 30
7. Results & Discussion 36
8. Conclusion 40
9. Future Scope 40
10. Limitation 41
11. References 42

**Abstract:-**

The traditional method of making judgments on the risk analysis in people suffering from coronary artery disease undergoing percutaneous coronary intervention (PCI) is done on the basis of limited clinical factors and medical report images. Decision-making is multiplexed in the modern health care system and rather is based on available data, structured understanding, and proper interpretation in the context of an individual patient. Machine learning has a major contribution in characterizing cardiovascular risk, outcome prediction, in biomarkers identification on available data of a huge population. Knowing the fact that a huge number of cardiac patients require PCI, there is no clear-cut outcome based on the application of Machine Learning to this specific group of patients to avail the advantage of Machine Learning in giving judgment on the analysis of precise prognostic endpoints on a large scale.

****

**Fig1**

**Introduction:-**

In recent years, percutaneous coronary intervention (PCI) has been a leading-edge innovation for the treatment of coronary artery disease. Risk stratification for diagnosis is crucial for patients and their personalized management who are undergoing PCI (percutaneous coronary intervention). An appraisal system for the long-term diagnosis of PCI patients may need to merge powerful all-rounder factors. Traditional prognostic risk assessment has limited strength to develop risk stratification. We designed a Machine Learning risk stratification tool that is able to assess and stratify the risk in various cases in patients prior to PCI. Through a comprehensive study, the best-performing machine learning model was random-forest which was used to predict and stratify patients in having different medical scores to generate a distinguished description of the model’s decisions and provide a precise result of individualized risk prediction and provide physicians an interpretation on the basis of key features of the given medical records.

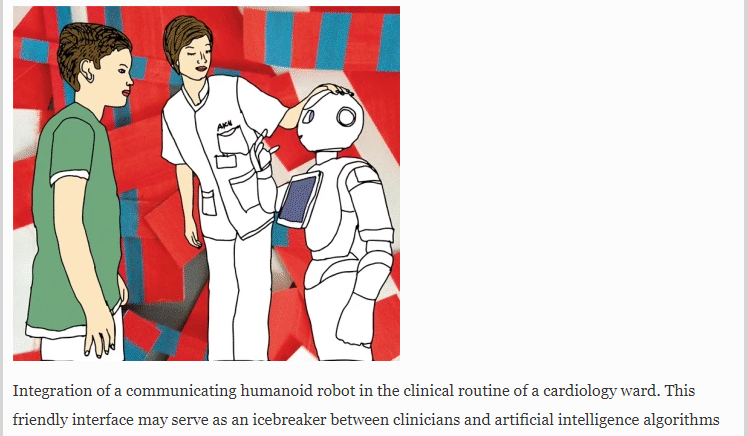


Fig2

**Literature Survey :-**

Here, we have studied the recent studies on the role of AI in cardiology and how AI can act as an interface between clinicians and treatment.

The first paper we referred was :

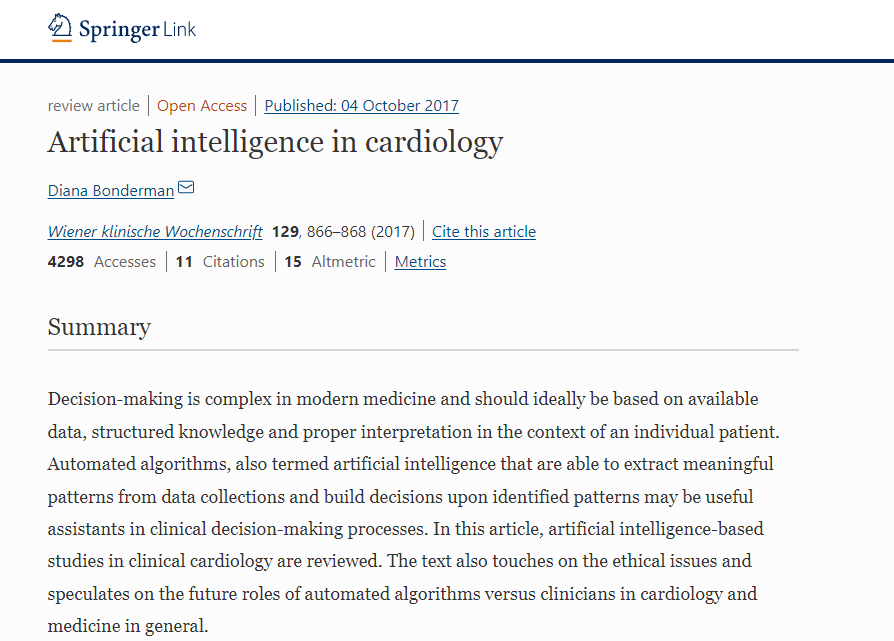


Fig3

The author of the study shows that the major advantageous feature of applying artificial intelligence in the health care system was: “There is no human error”. The implementations of AI in medical practice can deliver tools which can automatically stratify patients into specific prognosis classes and can save patients from invasive procedures, optimise time taking procedures, manpower and minimizes the grey zone inherent to human judgment.

Researchers had an opinion that a better phenotyping of affected patients might result in successful therapeutic strategies and identified pheno groups in 3 categories by specific differences in repolarization on their electrocardiograms. Group 1 has patients of young age having lower B‑type natriuretic peptide levels, group 2 comprises the highest prevalence of obesity and diabetes mellitus patients, and the group 3 category patients are the old age who have the most factors for chronic kidney disease. This was achieved by an unsupervised machine learning algorithm. This categorization has stimulated the field to pursue phenotype-specific research with medically highly relevant factors. They considered this successful application of AI in cardiology as a prefigure of a new cardiovascular medicine. They concluded that deep learning algorithms distinctly outperformed physicians in predicting prognosis and further treatment of patients with pulmonary hypertension.

The second paper we referred to was :

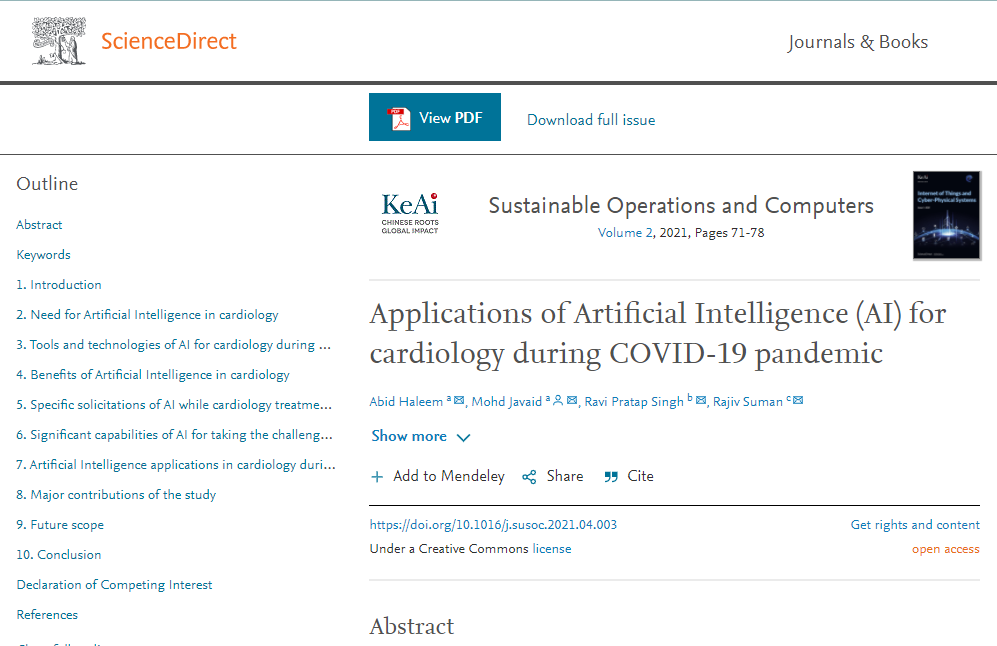


Fig4

We referred to the recent applications of AI during covid-19.We have studied a good number of research papers on Artificial Intelligence on cardiology during the COVID-19 pandemic to identify its significant benefits, applications, and future scope. AI uses artificial neural networks (ANN) to predict. In cardiology, it is used to predict the survival of a COVID-19 patient from heart failure.

During Covid 19 pandemic , artificial intelligence showed extreme positive effects on different health care areas. This paper tries to show the capabilities of artificial intelligence during covid in the field of cardiology. AI helps in measuring different functioning of the heart by providing its advanced technology. To bring out the applications , benefits and scope of AI in future we have studied a lot of papers based on artificial intelligence research on cardiology during covid 19. To predict something ANN means artificial neural networks are used by artificial intelligence. It is used to predict the lifespan of covid patients from heart failure. Artificial intelligence uses non common algorithms to predict diagnosis and treatments. Techniques like Machine learning and deep learning are used by AI to function. It's difficult to make sure of the results. Range of AI is large , during covid 19, it covered a large number of diseases , their roots of origin and prevention. And afterall a major application of AI is , providing technique in cardiology during covid 19. Among humans , a major cause of death is heart problems . Cardiovascular diseases multiply their number during covid pandemic. Treatment of these diseases are expensive so AI brings new technology to provide treatment at an affordable price. It is used in many ways like Identify the novel drug therapies. By cardiac based algorithms it predicts outcomes of covid 19. Many processes of treatment controlled by AI during covid 19.

To provide sophisticated data in cardiology digital technologies are used as monitoring devices. AI can take visual actions even during an ongoing pandemic , it's an intelligent system. To develop this type of intelligent system Machine learning is necessary. AI solves many heart related problems of covid patients with technology by treating complex problems. This uses many technologies like Artificial Neural Networks, Natural Language Processing, Support Vector Machines and Heuristics Analysis.

This artificial intelligence was introduced in 1956, but this was in very narrow range before covid. It helps us to learn new easy treatments and innovative ideas about cardiology and cardiovascular system. By applying artificial intelligence , the quality of education increased during covid 19. It is also used in reading complex heart surgery. AI continuously increases for reading accurate data than humans .

It affects positively by predicting and diagnosing cardiovascular health problems. AI overtake humans in reading accurate information and giving most probable results. It stimulates the intelligence of humans with computers. Congenital heart disease of covid 19 patients is assessed by AI . It is also helpful in relieving cardiologist burden. By using an automatic clinical decision system it helps to monitor the information alerts. Hence it is used to solve complicated problems. Work of cardiologists has decreased a lot, now a cardiologist can read complex heart problems digitally even at home. To analyse the medical data AI uses complex algorithms.

Applications of AI used for cardiovascular medicine process for clinical investigating the problem, computer science and in clinical practices. With AI commercial success is also linked . Not only cardio problems but also various climatic changes and environmental degradation are controlled by artificial intelligence. Also artificial intelligence helps in business to become more eco friendly and in creating sustainable product production and supply chain. It released a qualified model to develop profits from data .

Need for Artificial Intelligence in cardiology

There is a need for predicting accurate data from heart problems for covid 19 patients , and AI helps a lot in assuming accurately. In the cardiovascular system it predicts most accurate data and also in less time than any other technology. Also it helps in predicting heart failure of covid patients with accuracy and saves many lives. Applications of AI also used to get a clear image of the cardiovascular system. It also has the power to make improved diagnosis and prediction of heart related changes during covid 19. Main reason for increasing AI technology is that it works efficiently in less time.

Tools and techniques of AI

It works in smart and intelligent robots , efficient monitoring, cloud based data etc and helps a lot in solving health problems. It has smart and Powerful sensors for treating patients carefully and accurately. It also provides various advanced technology to handle cardio cases. Due to its effective philosophy, use of AI increased more during covid 19.

Benefits of Artificial Intelligence in cardiology

It improves connection between doctors, patients and administrative staff during covid 19. And it's most effective benefit was that it works effectively and productively. It also reduces time of investigation and giving results and also makes treatments affordable. It's fast prediction and easy treatments saved many lives , as doctors got enough time to treat if investigation of diseases was done as fast as possible. AI also helps in easy and early decision making by predicting accurate data. By using this technology, the efficiency of heart surgeons is increased . It's smart monitor and screen provide clear images of the cardiovascular system. AI reduces the risks of uncertain treatment. With the help of AI cardiologists increased their knowledge about the behaviour of patients. It improves protection and care of patients even at distance. It also helped a lot in research and development.

Capabilities of AI for taking challenges of covid 19

There are a lot of bad and dangerous impacts of covid 19 on the whole world. And simultaneously it directly affects the health care system and things related to it. And now in this situation there is a great requirement of analysing the data produced in this global pandemic. And AI with it's effective capabilities analyse the row data in a meaningful manner. It is capable of analyzing information about the transportation system of a country during this pandemic where many facilities do not work. It also helped in predicting the impact of global disease by measuring different regions. Also analysed the progress and reduction of covid in different areas. It measured effects of covid on different types of patients and provided better solutions and ideas for treatment . Its automatic tracking system proved a boon for all during covid time .

AI applications in cardiology during covid 19

It used mathematical algorithms to check the behaviour of patients. It increased performance and decision making power of cardiologists during covid by providing well processed tools and techniques. Applications like aortic valve implantation and stenosis of coronary arteries saved many lives . By analysing the heart anatomy of covid 19 patients it helps to treat the patient better. Its advanced technology improves the workflow and productive work helps to prevent infection.

By algorithms it speeds up the automation. It provided higher capabilities for analyzing the echocardiogram in a better way. MRI means magnetic resonance imaging helps in proper segmentation of cardiac ventricles. AI reduces the dependency and provides standard images of echocardiographic. It is also used in increasing performance and sorting out various issues easily. It helped in analysing blood pressure of patients and also detect the amount and need of blood in the body. It gives easy updation of blood in the body. By properly analysing the image of the heart it reduces chances of death due to heart failure. During covid it helped in managing conditions of patients. It's advanced technology reduces the risk of kidney failure and heart attacks.

AI plays a major role in managing, diagnosing and improving predictions and outcomes.

Major contribution and future scope

Main benefit of AI is predicting the accurate data and this reduces the death of patients. By analysing functions of the heart, showing clear images of the heart, providing information about conditions and risks, AI provides a positive impact in cardiology.

In the future this technology will provide innovative ideas and solutions to solve many complex problems related to cardiology. It will also be accessible for all due to its low price of treatment. And will improve the condition of easy and early decision making . This is our real time need and in future also we will follow AI rules to solve our problems. AI based machine learning will help cardiologists and doctors to overcome complex situations easily in the available time.

Artificial intelligence is used in many ways to perform different functions related to the heart during covid 19 . It is used to predict the outcomes and these accurate data helped doctors a lot. We can say this technology may not replace the doctors but we can proudly say that this technology helps doctors to work effectively. And to generate hypotheses in cardio life. AI is our all time requirement and by using this we can work properly because it provides easiness to human work. Smart work reduces the hard work of humans and saves time and life both. And not only in cardiology but also helped in analysing additional information about covid.

The third paper we referred to was :

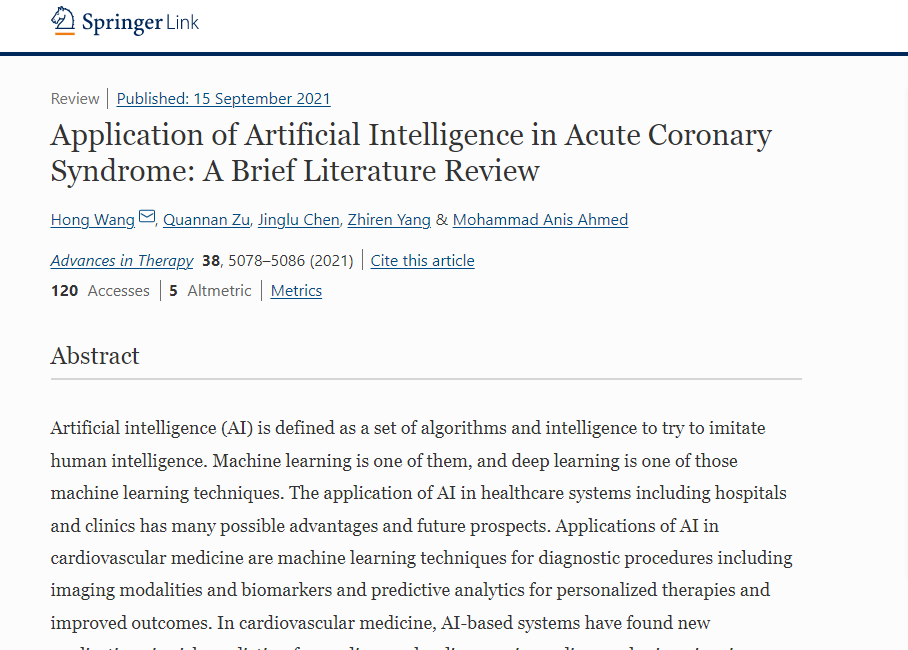


Fig5

We can define artificial intelligence as intelligence and algorithms which try to imitate the human mind. Machine learning is one of the techniques of artificial intelligence. And deep learning is one of machine learning. There are many advantages of artificial intelligence in healthcare services including clinics and hospitals . Moreover it has many advantages in different treatments and also has a secure future scope. In heart related medicine , applications of artificial intelligence are machine learning, used in the diagnosing process including image modalities and predictive data . Also artificial intelligence has found many new applications and techniques for risk prediction for heart related problems. Due to artificial intelligence and machine learning , we can predict a wide range of health related issues and factors including socioeconomic factors , comorbidities and a wide range of angiographic factors and agents related to it which was infused during angiography. Solving of Issues like stent malposition have been possible due to applications of artificial intelligence and machine learning. Moreover now machine learning can predict and identify the risk for patients with higher morbidity. It can also identify the risks due to acute coronary syndrome. And artificial intelligence made possible several potential benefits in victims of ACS. From starting to ending artificial intelligence can predict and identify the treatment of ACS and to prevent the adverse effects of ACS . Till now artificial intelligence and its applications have made an essential place in the clinical and medical world. AI is also used in interventional cardiology for management and treatment of patients with ACS. This paper will focus on reviews about applications of AI in ACS.

The fourth paper we studied was :

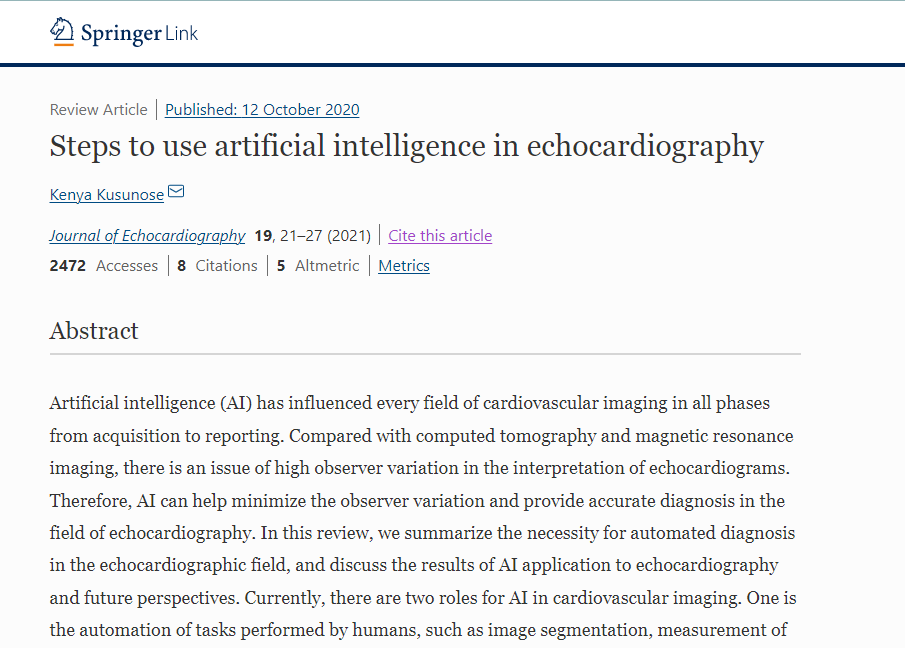


Fig6

During Covid 19 and even before it, artificial intelligence has influenced every field related to the heart . There are many observer variation issues in interpretation of cardiac systems but due to AI these variations are minimised. And this helps to prevent risk danger. And now cardiologists can get accurate information about echocardiography. In this review we will discuss the importance of automatic diagnosis in echocardiography. There are two main roles of artificial intelligence in cardiovascular imaging. The very first is the automation of those tasks which are performed by humans like measurement of cardiac structure and image segmentation. And the second one is discovering important insights of clinical issues. AI had reached a very high level during covid and even in future too there is surety of increasing applications of artificial intelligence.

If we define artificial intelligence,it is a process of eliminating and performing human tasks by computational program. For example pattern recognition by mimicking human thoughts. Artificial intelligence and its programs were introduced in the 1970s. But it was only used in gaming and social media. Around the 1980s machine learning was introduced and mainly used for robotics for a long time. But now it is relevant for medical use. It can work as a precursor to AI. Now computers provide accurate judgement by interpreting images and by giving control rules. After 2012 , along with AI and machine learning, deep learning was also introduced. And this helps to predict more accurate results. In the last 10 years machine learning and deep learning has developed a lot of new applications . Artificial intelligence covered a large area of medical science. For example by computing tomography and magnetic resonance of lumen diameter. Artificial intelligence also helped in minimizing the variations and providing accurate data for better results. AI in the COVID 19 era, there are many applications which were introduced in covid era . With development of applications for echocardiography, in routine examinations the number of parameters of echocardiography has increased. Now by using these applications we have become more complex and can solve more complex problems. Portable devices covered a large area of physicians and cardiologists, who are working from home by using these devices. Due to coronavirus, many industries and lifestyles were affected a lot. But many industries had a great profit also like during covid era people increased their dependence on telecommunications. Not only this part but also affects the medical world. This wave of infection affected cardiologists and their working and also affected the priority of patients. Now people prefer to check up at home instead of visiting hospitals. Before covid era patients were moved to hospitals and underwent an echocardiogram check up by experts or cardiologists. But now personal protection is a must , still many patients require a quick and accurate diagnosis and this is reduced to limit possible exposure. By taking only images instead of doing echocardiography study we are missing many important findings. This can lead to poor outcomes of patients but by AI we can reduce these problems. We should know the main difference between deep learning and conventional machines to understand the role of deep learning in echocardiography. And the main difference is that in deep learning multilayer neural networks are being used. This improves the quality of results mainly in visual objects. When we learn by using artificial intelligence we can divide it into three categories . First is supervised learning, second is unsupervised learning and third and last is reinforcement learning. Supervised learning is learning by finding the training data with accurate answers. Many images with labels should be included when we make a new model in the echocardiographic field. Unsupervised learning is a learning method without any labels. A cat could recognise without manual training was the famous example of AI created by Google. This learning process is similar to human learning. Here we will focus more on the real type of learning that is supervised learning. Now AI has gained 80 to 90 percent accuracy , and it has become possible to create accurate models without a lot of images.

Steps of echocardiographic AI

In management of cardiovascular diseases and their diagnosis echocardiography is useful . For good medical practices accurate studies of echocardiography is necessary. And this requires an adequate amount of images as well as their proper interpretation. This process involves four steps. And while developing applications these steps are necessary to divide into two separate tasks.

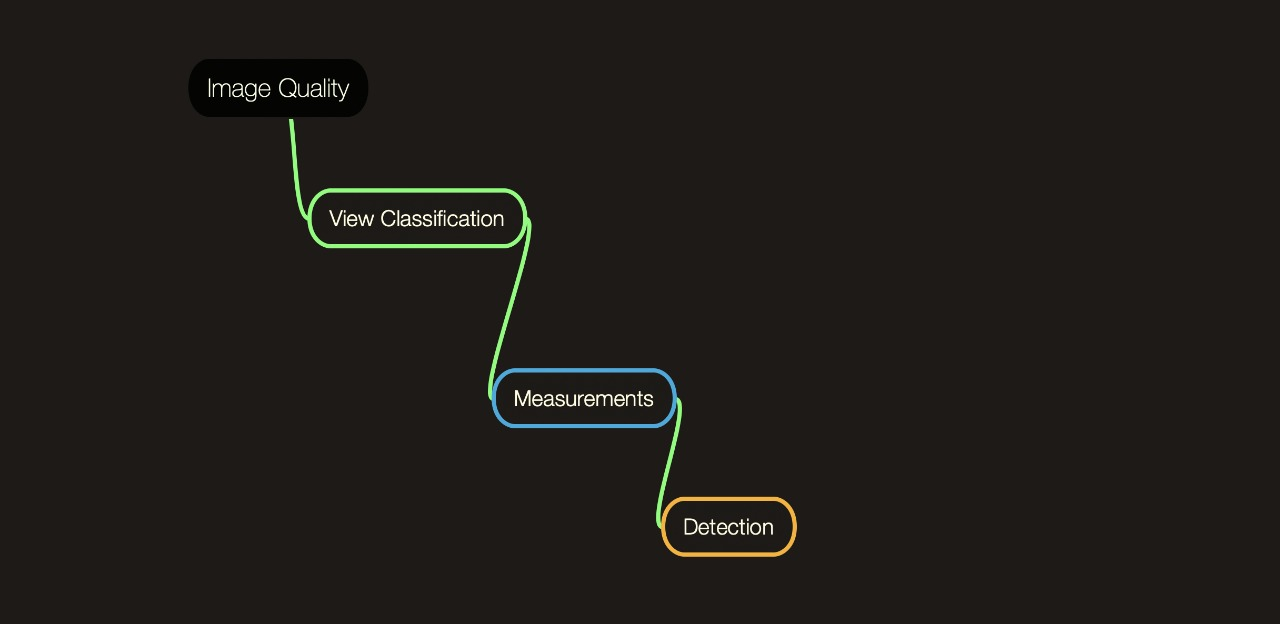


Fig7

1st step : maintaining quality of image

This step is the evaluation of the quality of the image in applications of AI . Among humans , their experience obtains correct images. If quality of images are poor these must be examined by inexperienced examiners. Recently research shows that image quality in AI was excellent . Recently in America around February 2020, FDA ( food and drug administration ) approved the first device based on AI based echocardiography . This guides the users to find a better image.

2nd step: view classification

This includes the view of segmentation of cardiovascular structures. Images of echocardiography need many types of recordings because of its complexity in the cardiac system. This classification can be used in detection of appropriate views. Many researchers found a good accuracy for view classification models. Accuracy attained 91 to 94 percent for 15 view classification. A new view classification based on neural networks having 17000 images was reported recently. And this contained 1.9 percent of mislabeled images. This accuracy rate was good enough for creating a feasible prediction model.

3rd step : measurements

After having a good quality of image , the next required step is to measure and quantify the morphological structure of this image. For an example we can take different frameworks between deep learning and conventional machines for ejection fraction. After processing the image , to get extraction of features if the target region human scientists focus on manually configuration of a point. This method is based on deep learning and it approaches fully automated echocardiographic data. A mean percentage error of 10 percent was reported by their model for EF chamber 2 of aoic and 20 percent for apical 4 chamber.

4th step: detection of abnormalities

This is the last step but not the least. This involves detection of abnormalities in the model. Most common and important detection is regional wall motion abnormalities; it is an important evaluation in echocardiography in which a patient's complaint about chest pain is a class in many cardiovascular guidelines. To remove these abnormalities a machine learning model was introduced , to identify the RWMA these were developed. By using deep learning algorithms RWMA can be decreased.

There are presently two main roles of artificial intelligence in cardiovascular imaging. First is automation of human tasks and the second one is image segmentation. Most applications were mainly focused on automation of hunan tasks. For getting accurate measurement nany applications were developed. Artificial intelligence can be used in everyday medical practices and become valuable for saving many lives and for dealing with cardiovascular health problems.

The fifth paper we referred gave us the direction to decide the topic of the project :

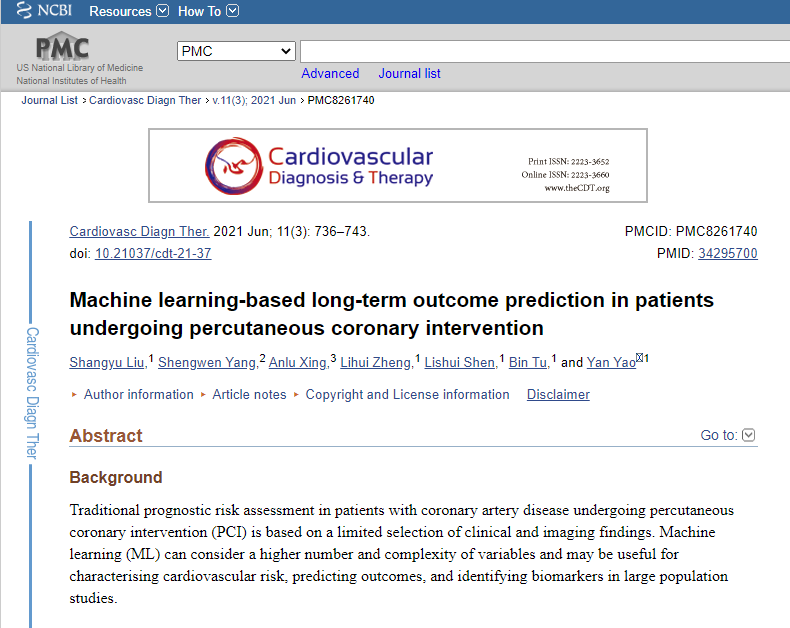


Fig8

PCI means percutaneous coronary intervention, and risk assessment of victims of PCI with coronary artery disease is based on a limited selection of medical findings . Machine learning can be used in treatment of various complex cardiovascular health issues. It is also used in predicting risks of heart disease.

Around 9680 patients underwent PCI who were suffering from coronary artery disease , between January 2013 and December 2013 at institutions in the PCI area. 6 different ML models were trained by different selected clinical features. Many gradient boosters were applied to increase the performance of models.

According to a survey held in the last five years , it was concluded that 467 patients died . And around 87 baseline of risks were used to train ML models. Best performance was given by a random forest model. That was also called RF PCI. Among 11 laboratory measures, the top 15 features of PCI were prognosis.

Long term prediction of mortality due to coronary artery disease was improved by ML models, before PCI. And the best performer was the RF model, it provided satisfaction to the researchers.

PCI treatment was a revolutionary innovation in past years. Risk satisfaction for prognosis is important for individuals in the management of coronary artery heart patients. There is a great need to integrate powerful mukti features. GRACE means global registry of acute coronary events evaluate the short term prognosis of coronary artery heart patients. thrombolysis in myocardial infarction was also considered. Now let's discuss machine learning . It is a field of computer science , in this complex rules are applied to solve complex problems and multidimensional technologies are used to work effectively.

Around 9.5 thousand coronary artery heart patients underwent PCI and 87 risk factors were collected from different persons. And to analyse the results many groups were formed to categorize the patients. Like physical characteristics , laboratory measures, electrophysiological results etc.

And hypertension was also considered a history of many patients . Diabetes patients, blood pressure patients were also considered while doing the survey. And moreover, smoking history was recorded within the last 3 months. Many certified sonographers use two dimensional echocardiography to consider the parameters of echocardiography.

To select the appropriate features, information gain attribute ranking was used. And if we define the information gain we can say it is a measure of effectiveness and wellness of an attribute while classifying the training data. It also reflects the additional information about the class provided by the features at which entropy of a class decreases.

During analysis, all six ML performances were compared in predicting all causes of death in patients of coronary heart disease. To evaluate the performance of all six models, all 10 folds of cross validation were used. This process was repeated one after one around 10 times to get unique results.

By using 23 versions of SPSS statistics, all analyses were performed . The Python foundation was also used for analysis. All the data were presented as standard deviations and mean deviations for continuous variables. And this counted the percentages for categorical variables. For normal data, results were also as mean and median .

There were 9680 patients with heart disease related to coronary disease who ever underwent PCI. Mean age for that cohort was 58.5 years. In that survey of models there were 76.92 males and 64.61 percent had a history of hypertension. And moreover around 30 percent of patients had diabetes.

Ranking criteria for this study gave information about ML models and according to it there were 15 top-ranked features in RF PCI and 11 of them were top features for all causes of mortality in the previous 5 years.

Among all models and classifiers , RF means the random forest model was the best performer. Average AUC shown by the RF model was over 0.71 , which can be used to find and predict long term prognosis of PCI patients. Main key for coronary heart problems and their treatment and diagnosis is prognosis assessment. We can say it is the best intervention method if we consider different risk stratification. Predictive and profound scores of heart problems related to coronary artery disease have been widely used in risk analysis before revascularization . Many factors that are considered independent have limited predictive value for disease . Machine learning is also a subset of artificial intelligence , which has the ability to learn independently to make accurate and effective predictions. This technique has developed rapidly in the last few years. It has been used with artificial intelligence for medical purposes.

In this whole research we developed an ML based model to find the medical outcomes of PCI undergoing heart patients. Performance of the random forest model was more satisfactory than others, in providing useful information and results for any predictive model for PCI clients for long term outcomes. Ethical statement stated that the accountability for all aspects of the work was of the authors , no matter what it was for. Whether it is related to integrity or accuracy or any part related to approximation.access article. This article was distributed in accordance with creative commons attribution non-commercial international license. This licence permits the replication of non commercial articles with strict proviso.

Model Description :-

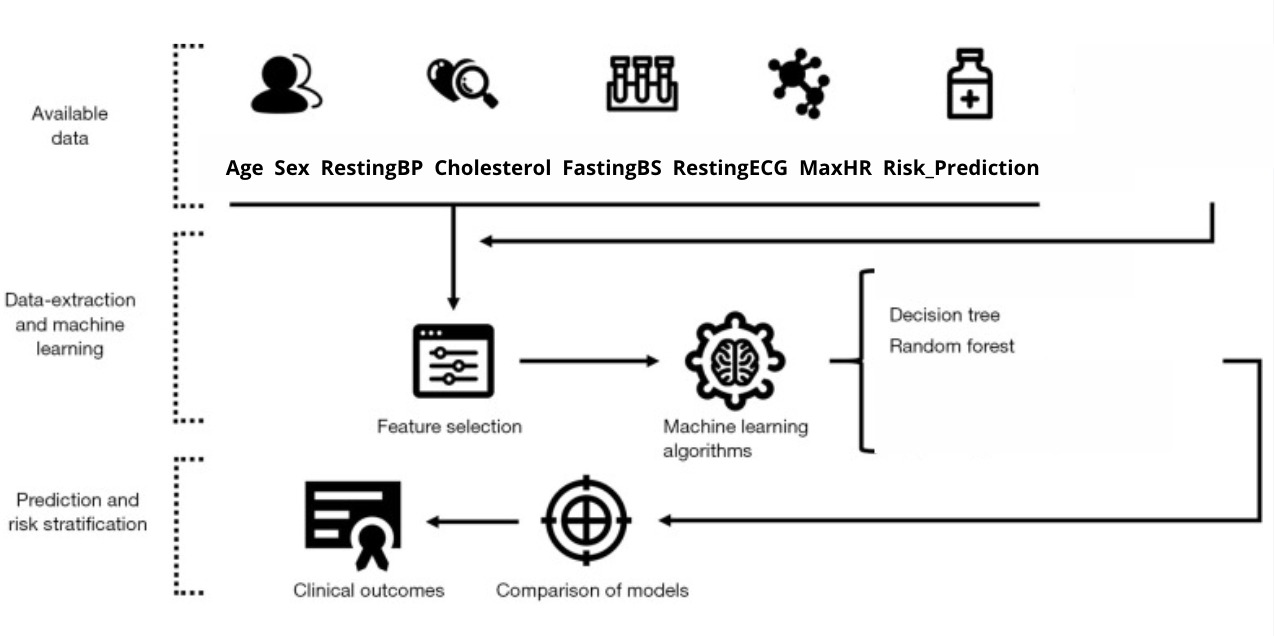


Fig9

With the help of all the research papers we referred to, we came to know that decision tree and random forest algorithms outperforms by giving the maximum accuracy. Decision tree is the most influential and simplified algorithm for classification and prediction. It is a schema-like tree structure, where each internal step indicates a test on a feature, each branch signifies an outcome of the test, and each leaf node holds a decision.

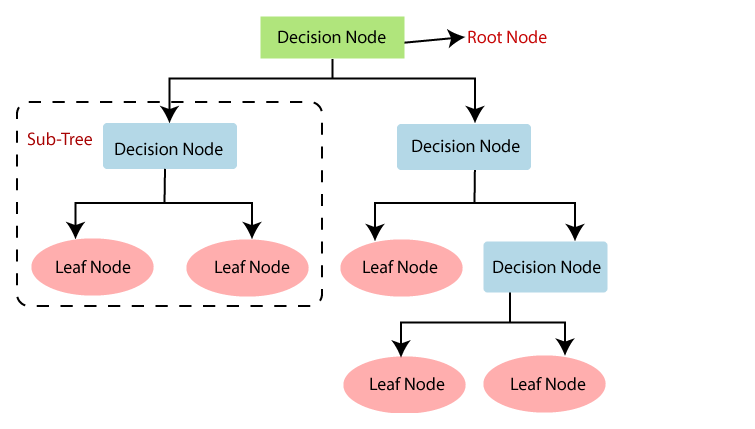


Fig10

It is a supervised learning algorithm, we made the algorithm learn classification using training data and predict using testing data.

Diagram

To learn classification, data must know the best feature at each step while building a Decision tree, for that we must find which feature gives the most valuable information using the concept of information gain. Information Gain helps in calculating the decrease in the entropy and finds how well an attribute categorises the target classes. The attribute which has the maximum value of Information Gain is considered as the best feature. To calculate the information gain, we must know the entropy of each feature. Entropy is basically a measure of dissimilarity in the target variable in a dataset. In the case of binary classification,

if entropy is 0, then all values in the target variable are the same(either all positives or all negatives)

if entropy is 1, then the target variable has equal number positive and negative values.

Entropy is calculated as:

|  |
| --- |

S -> entropy

n -> total number of classes in the target column ,in our case n = 2 i.e 1(risky) and 0(not risky)

pᵢ -> probability of class ‘i’ or the ratio of “number of rows with class i in the target column” to the “total number of rows” in the dataset.

Information Gain for a feature column A is calculated as:

|  |
| --- |

Sᵥ -> set of rows in S for which the feature column A has value v

|Sᵥ| -> number of rows in Sᵥ

|S| -> number of rows in S

Another factor for a model to consider while learning classification is the purity while creating a decision tree. This can be measured using the Gini index. Low Gini index feature must be preferred as compared to a high Gini Index value feature. This index is helpful in creating binary splits.

Gini index can be calculated as :

|  |
| --- |

The benefits of using this model were,

Foremost reason is that it is a simplified algorithm and easy to understand and very relatable with the decision in real-life.

It is an apt algorithm for us as our problem statement is based on decision making and after considering all the parameters we get outcomes

The most tedious task for performing any algorithm is data cleaning, and this algorithm reduces our efforts as it has less requirement as compared to other algorithms.

As every solution has a problem and that problem can also be resolved. The same happens in our project. We have a huge dataset and have multi-classification of each feature. We get a lengthy and not so easy to read decision tree, as the resultant decision has lots of layers. The multi-classification of each feature made it computationally complex and it may have overfitting issues too.

An overfitting examination is a method for finding how and when a particular machine learning model is overfitting on a given dataset. It is a technique that can help us explore the learning dynamics of a machine learning model. This can be performed by analysing the behavior of the model in the midst of a single run for models like neural networks which get fit on the training data incrementally. A graph of performance of the model on the training and testing dataset can be evaluated at every point while training the data and later plots can be generated using these points. This graph is known as a learning curve plot in which one curve shows the performance of the model on the training set and one curve shows the performance of the model on the test set for every increment of learning.

The research papers came into the picture again and we got the solution to the new problem as well and i.e Random Forest Algorithm.

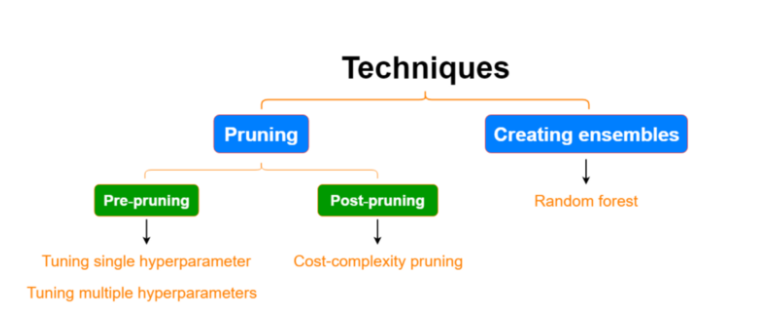


FIg11

Random Forest

Random forest is a pliant, easy to understand [machine learning algorithm](https://builtin.com/data-science/introduction-to-machine-learning) that gives good results most of the time. It is one of the most practiced algorithms, because of its flexibility and multifariousness i.e it is used for both classification and regression problems.

It works on ensemble learning. Ensemble learning is a process that integrates multiple classifiers or models to provide solutions to complex problems and to make predictions rather than an individual model.

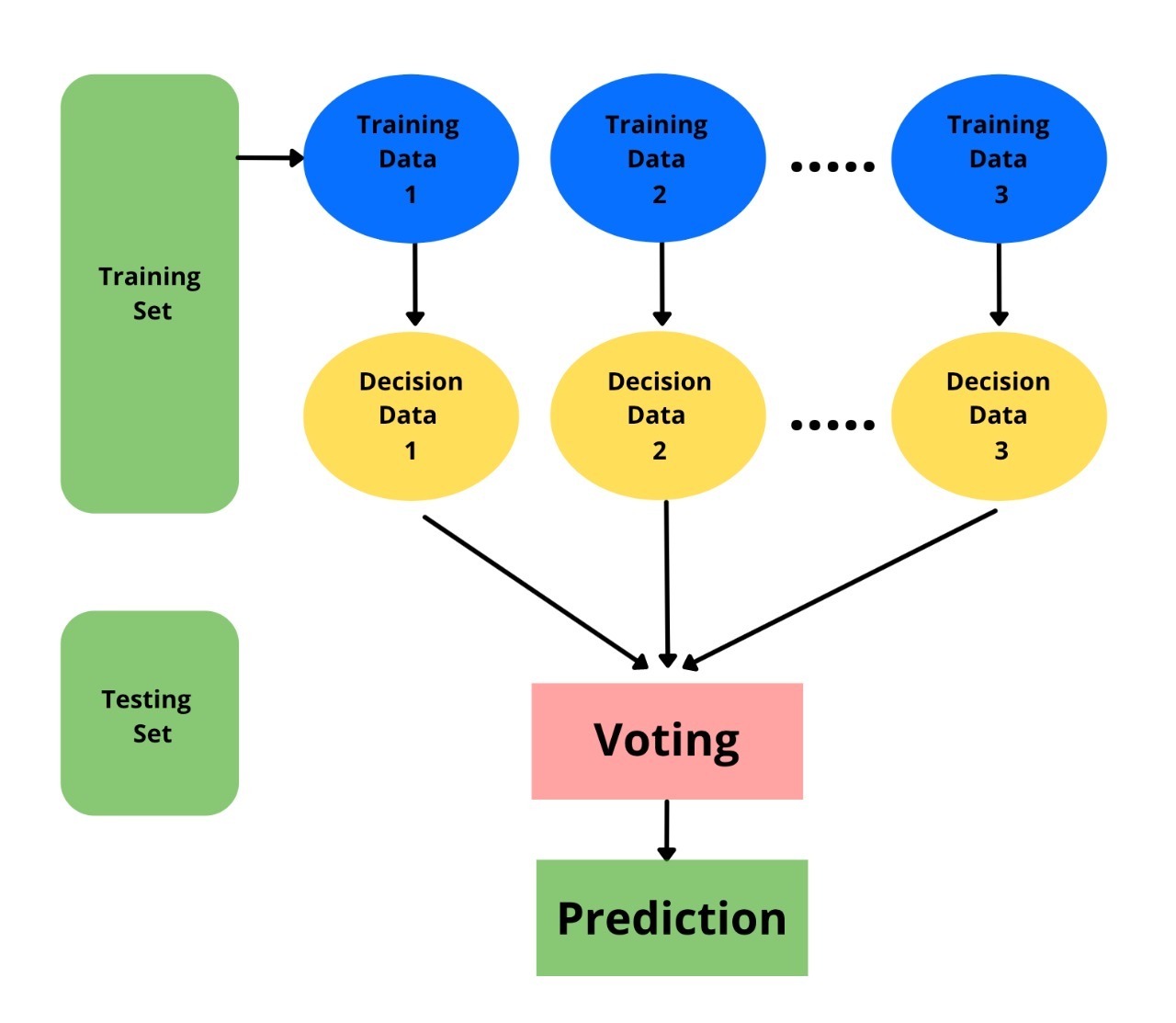


Fig12

A random forest algorithm comprises many decision trees. The ‘forest’ created by the random forest algorithm is trained using bagging or bootstrap aggregating techniques. Bagging is a method that improves the performance of machine learning algorithms. It is also called Bootstrap Aggregation, selects a random set from the data. Each model created from the samples from the given Data with substitution is called row sampling. Row sampling with substitution is called bootstrap. Further, individual model’s training takes place independently which gives the results. The end result is based on maximum voting after integrating the outcomes of all models. This process of combining all the results and calculating output based on maximum voting is called aggregation.

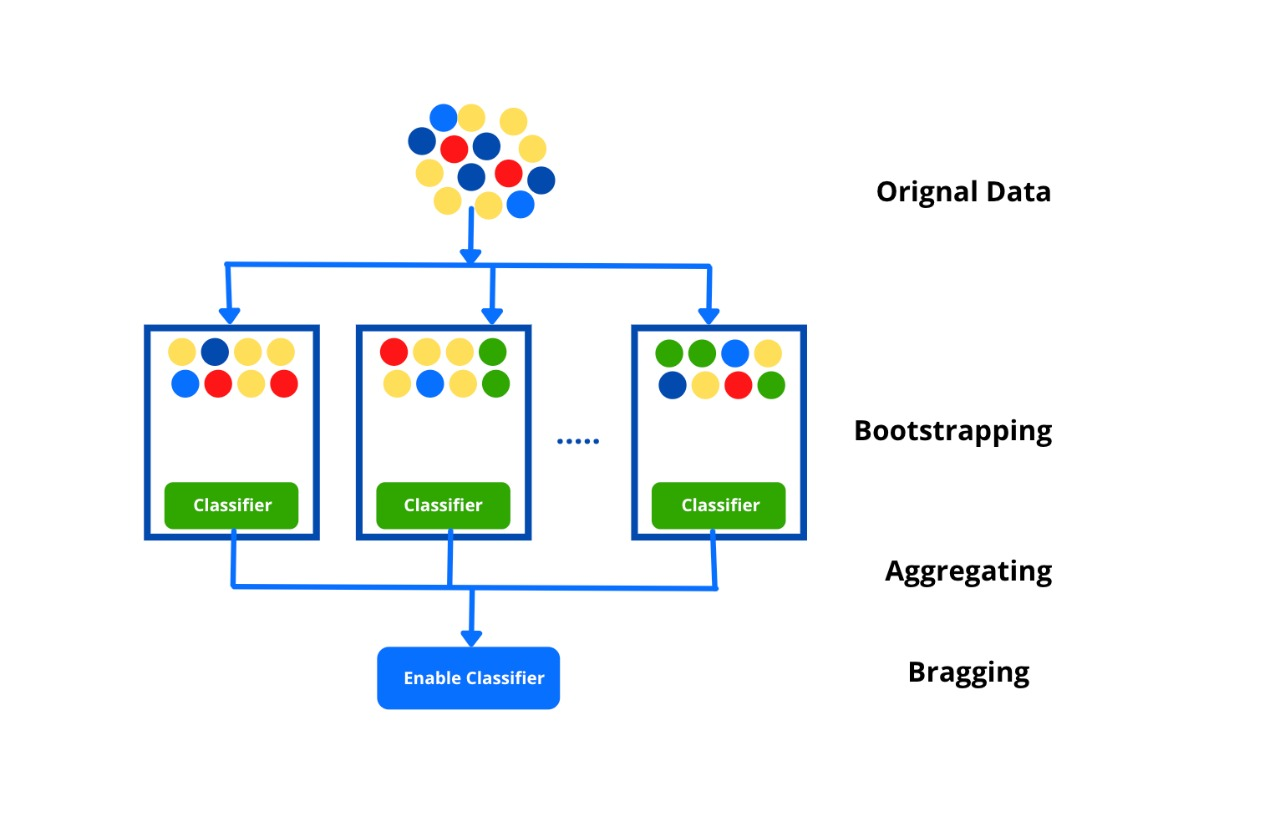


Fig13

Random forest algorithm establishes the result based on the predictions of the decision trees. It calculates the result by averaging the outputs from multiple decision trees. Precision of the results increases on increasing the number of trees. A random forest overcomes the disadvantages of a decision tree algorithm. It decreases the overfitting of datasets and improves precision. It makes predictions without the need of many configurations in packages.

Data Description:-

Very little research has been done in this domain, so it was quite tedious to find the appropriate dataset. The dataset consists of the following medical records of 918 patients for whom physicians want to analyse whether the patient may face risk or not prior to PCI.

| **Age** | **Sex** | **RestingBP** | **Cholesterol** | **FastingBS** | **RestingECG** | **MaxHR** | **Risk\_Prediction** |
| --- | --- | --- | --- | --- | --- | --- | --- |

These medical factors have further classification which helps in making decisions and generating decision trees.

| **Feature** | **Description Of Feature** |
| --- | --- |
| Age | Ranges 28-77 |
| Sex | Male |
| Female |
| RestingECG (Electrocardiography) | Normal |
| ST |
| LVH |
| FastingBS (fasting blood sugar) | Yes |
| No |
| RestingBP (Blood Pressure) | Ranges 0-200 |
| Cholesterol | Ranges 0-603 |
| MaxHR (Maximum Heart Rate) | Ranges 66-200 |

Data Preprocessing:-

Data preprocessing is a method of making the raw data appropriate for a machine learning model. It is the foremost and vital step while building a machine learning model. When building a machine learning model, it does not always happen that we come across neat and well formatted data. And operating with data, it is necessary to clean it up and put it in a well formatted way. So for this, we do data preprocessing tasks. Pre-processing infers the transformations implemented to the data before using it for the algorithm. Data Preprocessing is a method to transform the raw data into a formatted data set. In simple words, whenever the data is collected from various sources it is received in raw format which is not formatted for the analysis. When it comes to designing a Machine Learning model, data preprocessing is the foremost step marking the commencement of the process. Usually, real-world data is insufficient, conflicting, imperfect (contains errors or outliers), and most of the time lacks particular feature values. Here data preprocessing comes into the picture– it helps in data cleaning, data formatting, and data organizing of the raw data, as a result making it ready-to-use for Machine Learning models. Following are various steps of data preprocessing in machine learning we followed:

1. Importing Libraries : To perform data preprocessing using Python, we require importing some already defined Python libraries. These inbuilt libraries perform some particular jobs. There are three libraries that we used for data preprocessing :

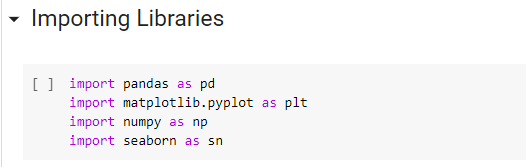


Fig14

1. Importing Dataset : We imported the datasets which we have collected from kaggle for our machine learning project. We integrate the dataset with google colaboratory file.

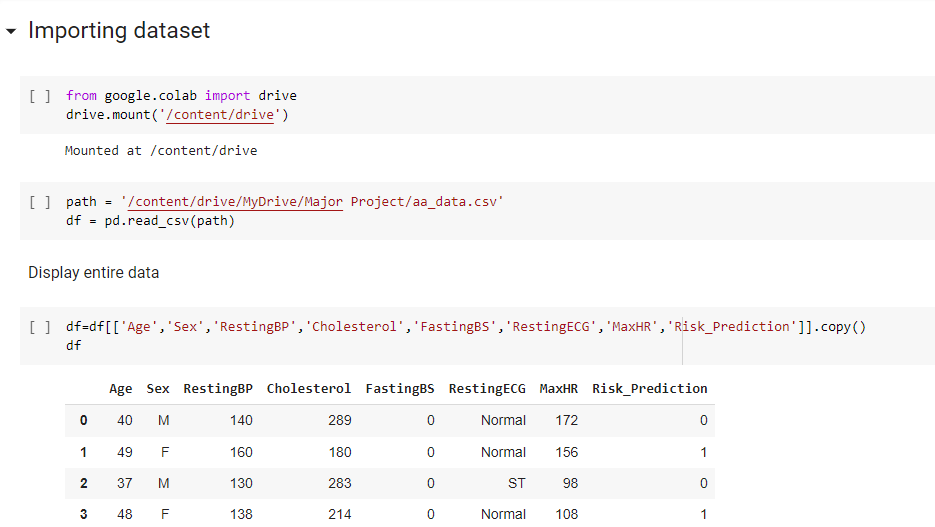


Fig15

1. Handling Missing Data: The next step we performed was handling missing data in the dataset. If the dataset contains some missing data, then it may create a huge problem for our machine learning model. Hence it is necessary to handle missing values present in the dataset. In our case, we didn’t face this issue.

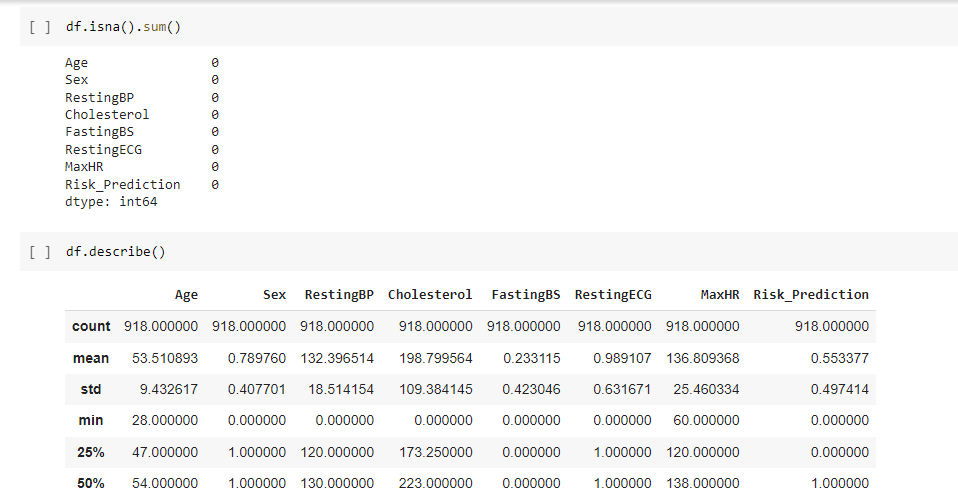


Fig16

1. Encoding Categorical Data: Categorical data is the data which has some categories like, in our dataset, there are two categorical variables, Sex and RestingECG. Since the ML model completely works on numerical values, and if the dataset contains a categorical variable, then it must be in numerical form otherwise it may create trouble while creating the model. So it is required to encode these categorical variables into numbers.First, we converted the Sex variables into categorical data followed by the RestingECG variable. For this, we used the LabelEncoder**()** class from the preprocessing library.

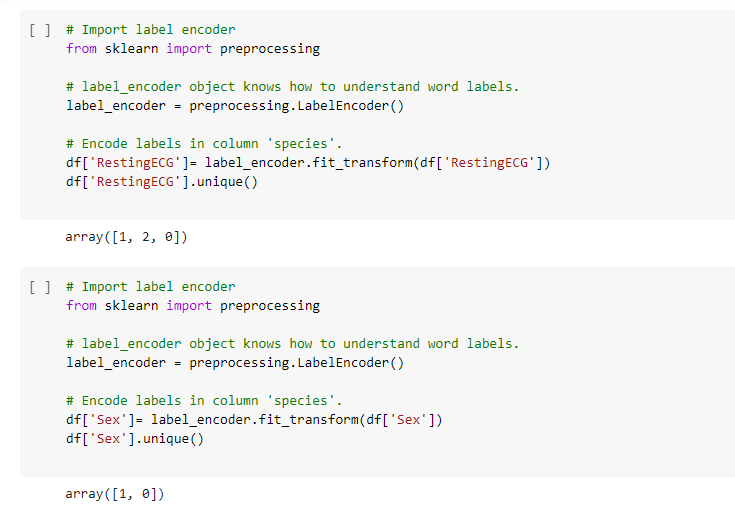


Fig16

1. Splitting Data into a training and testing set: We divided our dataset into a training set and test set. This is one of the vital steps in data preprocessing. By doing so, the performance of our machine learning model enhanced. For instance, if we train our machine learning model with a dataset and test the model with a completely different dataset. Then, it will be difficult for the ML model to grasp the correlations between the features. On the other hand, If we train the model and its training accuracy is good, and while testing the model we provide a new dataset to it, then it will decrease the performance. Therefore, we always build a machine learning model which executes well with the training as well as with the test dataset. Here, we split these datasets:

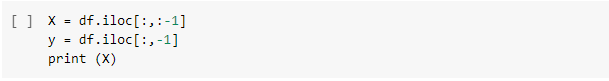




Fig17

Training and testing data

1. Decision Tree Model

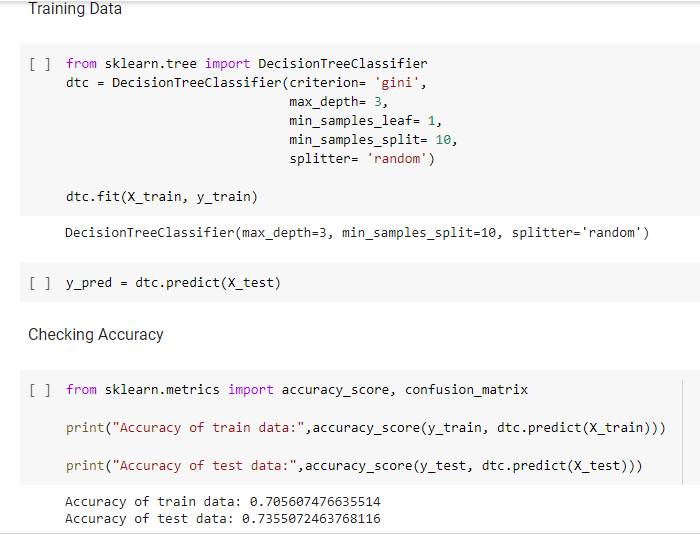


Fig18

1. Random Forest Model

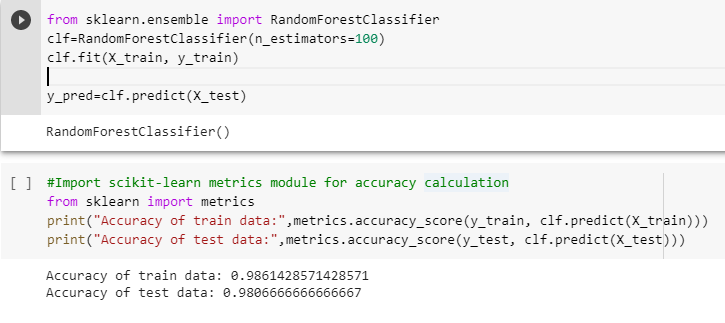


Fig19

Results:-

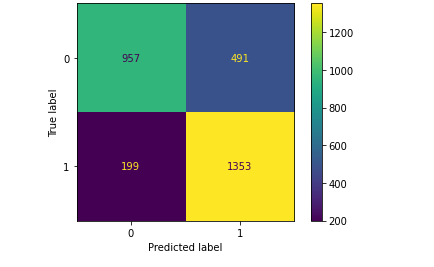


Fig20

Confusion matrix of decision tree

This is the resultant Confusion matrix of the Decision Tree which clearly gives bad results in case of false-negative i.e if the patient may have risk after undergoing PCI, the model is predicting that he will have no risk and false-positive cases i.e if the patient has no risk after undergoing PCI, the model is predicting that he will have risk. These results may guide physicians in the wrong direction of treatment. The reason behind wrong predictions can be the overfitting of the model.

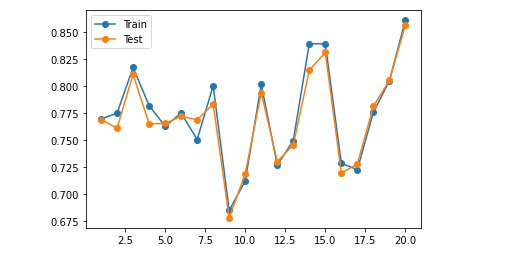


Fig21

Overfitting in Decision tree Results

On visualizing the overfitting in the model, we found that the reason we assumed was correct. To be more sure, we decided to review the performance of the algorithm using a performance matrix. Performance matrix helps us decide the quality of prepared model, it shows our test dataset is distributed across four different sections called False positive(FP), False Negative(FN), True Positive(TN) and True Negative(TN).The following values help us calculate the accuracy, precisions and F1 score of our model as shown below.

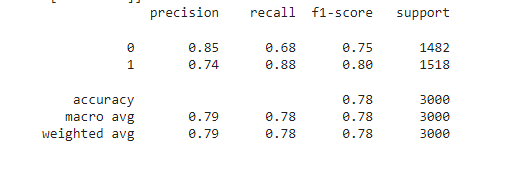


Fig22

Confusion matrix report

Calculation of Confusion Matrix Report

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

Precision = TP/(TP+FP)

Recall = TP/(TP+FN)

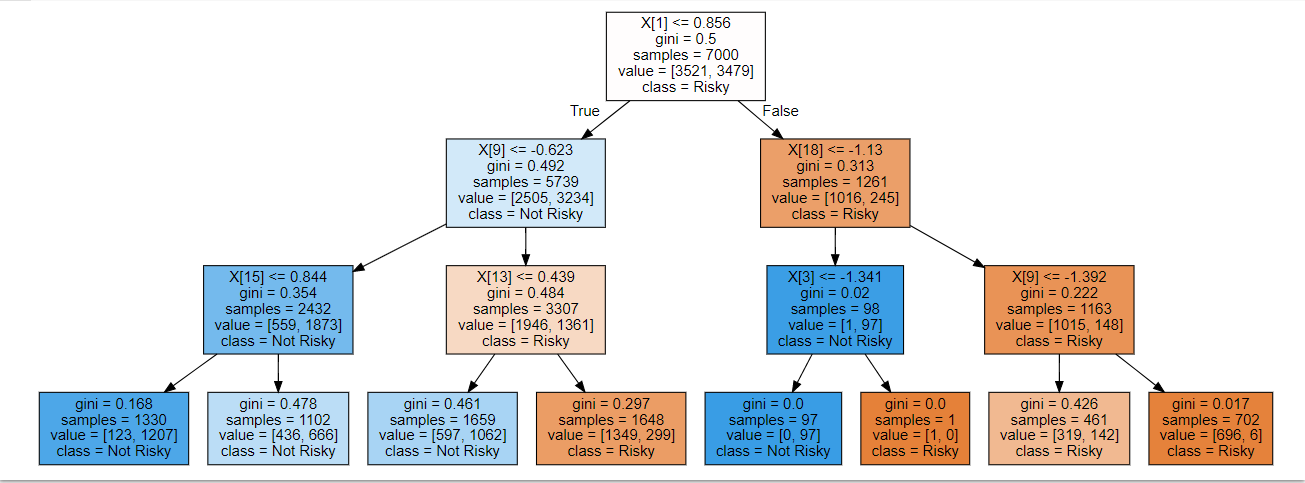
Finally we visualize the decision tree generated using the decision tree classifier. 

Fig23

Decision tree

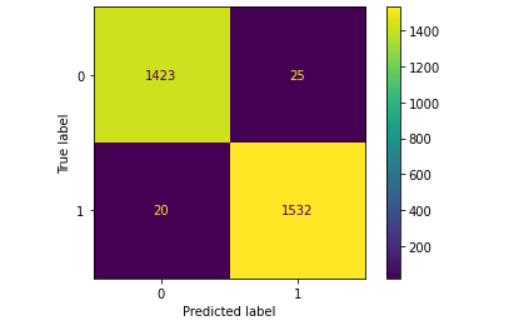


Fig24

Random Forest Confusion Matrix

After designing the Random forest algorithm, The above confusion matrix generated much better outcomes as compared to decision tree model.The count values of True positive and and true negative increases whereas the count of false negative and false positive decreases with a good difference from that of decision tree model. We again checked the performance of the model using the performance matrix and found way better results as compared to the decision tree classifier model. The accuracy increased from 78% to 98% which is a huge difference.

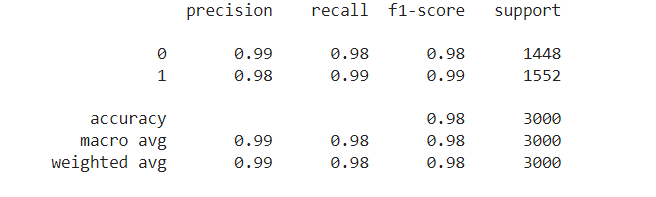


Fig25

Confusion matrix report

Calculation of Confusion Matrix Report

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

Precision = TP/(TP+FP)

Recall = TP/(TP+FN)

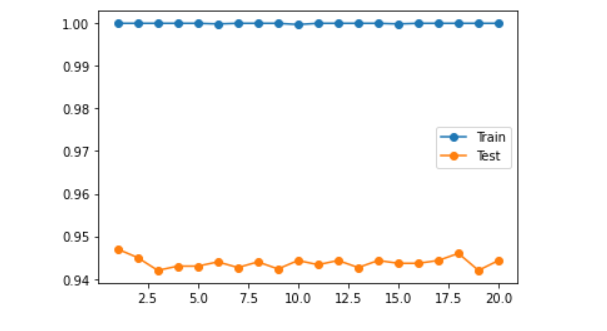


Fig26

Overcome overfitting limitation of decision tree

Even without tuning any of the hyperparameters, the random forest model performs well compared to the decision tree model.

**Future Work**

* **Implementation of GANs** (Generative Adversarial Networks)

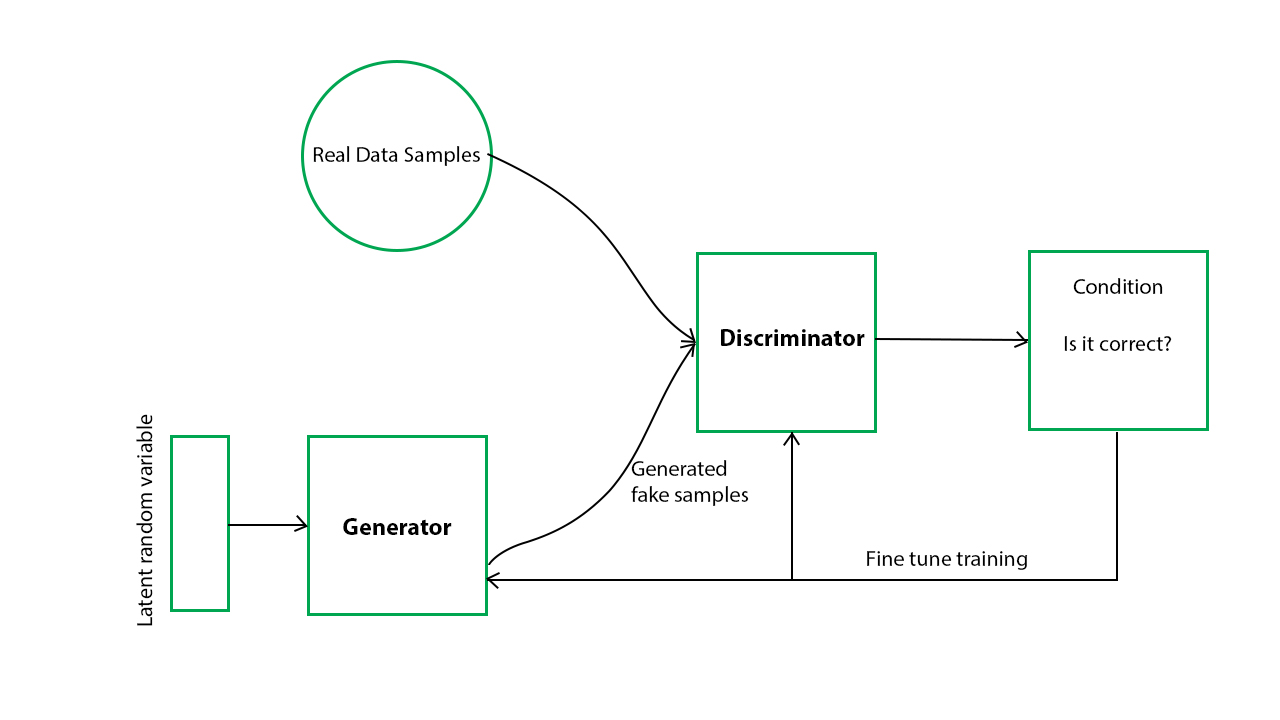


FIg27

**Generative antagonistic organizations (GANs)** are algorithmic designs that do the utilization two neural organizations, setting one in opposition to the next to produce new, manufactured cases of information that can be mistaken for genuine information

* **Improving Model’s Accuracy**

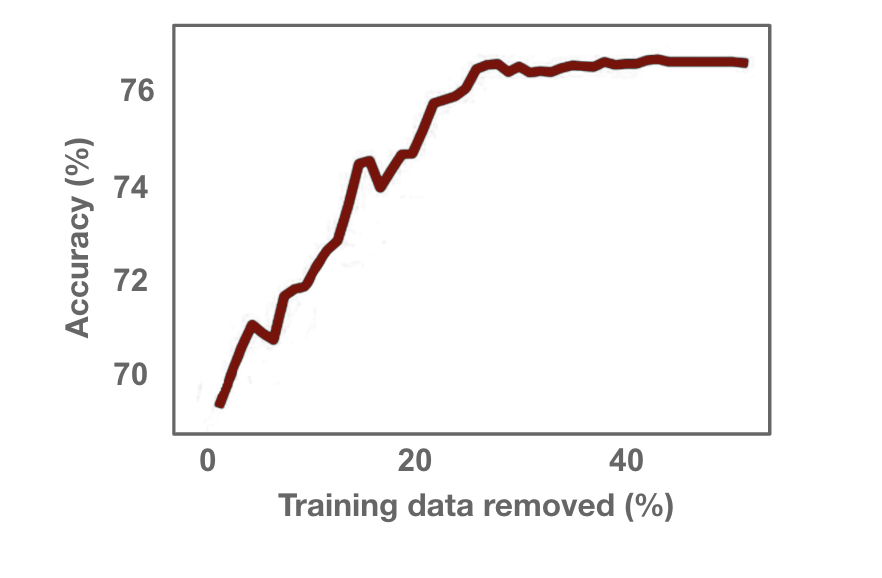


Fig28

We can improve the accuracy of this model by trying and testing different algorithms present in the current market and if still there is some scope left then we can create our own algo to make this model work more precisely and seamlessly.

**limitations** :   
  
\* The the data which is preset in our dataset is not enough to predict any medical result or in other words we can say it can be a immature decision in some multi complication scenarios

References

1. <https://link.springer.com/article/10.1007/s00508-017-1275-y>

# Artificial intelligence in cardiology

1. <https://link.springer.com/article/10.1007/s12574-020-00496-4#citeas>
2. Steps to use artificial intelligence in echocardiography
3. <https://onlinelibrary.wiley.com/doi/full/10.1002/emp2.12205>
   1. identify top clinical variables and constructed a risk stratification score to predict mortality in COVID-19 patients
4. <https://obgyn.onlinelibrary.wiley.com/doi/full/10.1002/pd.5892>
   1. AI will form at least some part of routine fetal cardiac care in the near future. We envisage a future where AI works in tandem with skilled clinicians to optimize performance.
5. <https://www.sciencedirect.com/science/article/pii/S2666412721000155>
   1. Applications of AI in cardiology
6. <https://drive.google.com/file/d/1rVGbNYw-xlRm37OBT0ERwB0pqtgzwkqE/view>
   1. 3-year all-cause mortality in patients with heart failure (HF) caused by coronary heart disease (CHD).
7. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8261740/>
8. develop and validate different ML-based models to predict five-year all-cause mortality in patients with coronary heart disease prior to PCI and
9. identify the important predictors for clinical outcomes in a large population.

8. <https://innovations.bmj.com/content/7/3/564#article-bottom>

1. develop and validate a set of artificial neural networks (ANN) models to predict five adverse outcomes after PCI—AKI, bleeding, stroke, death and any adverse outcome.

9. <https://link.springer.com/article/10.1007/s12325-021-01908-2>

# Application of Artificial Intelligence in Acute Coronary Syndrome: A Brief Literature Review

10. <https://drive.google.com/file/d/121uH4U9Zkvnpd_ZxDQ8O-TM_xrgT3Zve/view?usp=sharing>

assessed the ability of an artificial intelligence algorithm (ECGio) to predict the presence, location, and severity of coronary artery lesions in an unselected stable patient population referred for coronary angiography.