INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 06, June 2023, pp : 437-447

e-ISSN : 2583-1062

Impact Factor : 5.725

A SMART HEALTHCARE ALGORITHM FOR ANALYSIS AND PREDICTION OF HEART DISEASE USING ML

Nandan kishor[1]

[1]Master in Technology, Department of Computer Science and Engineering, Technocrats Institute of Technology, RGPV, Bhopal, India

DOI: https://www.doi.org/10.58257/IJPREMS31726

ABSTRACT

In the current decade heart disease is most common for all from 14-60 years of age. Now a days heart disease having many types of constraint to predict the information like hyper tension, Blood pressure increase, Blockage of Nervous system. So that the percentage of blockage must identify which surgery is good for the health of patient. In this paper we are trying to find the better method to predict and we also used algorithms for prediction. We use various algorithms for the prediction of blockage of nerves like Naïve Bayes; algorithm is analyzed on dataset based on hazard factors. Here we also try to use decision trees and grouping of algorithms for the calculation of heart disease based on the features. The results shown the comparison of various four algorithms of data mining which compare their accuracy of the result which one is better through graph.

Keywords: Decision tree, Data mining, Heart Disease Prediction, Naïve Bayes, K-means, Machine learning.

1. INTRODUCTION

Coronary illness is the sort of infection that can cause passing. Every year such a large number of people groups are biting the dust because of coronary illness. Coronary illness can be happened because of the debilitating of the heart muscle. Additionally, a cardiovascular breakdown can be depicted as the disappointment of the heart to siphon blood. Coronary illness is additionally called coronary course infection (CAD). Computer-aided design can be happened because of inadequate blood supply to veins. Coronary illness can be identified utilizing side effects like hypertension, chest torment, hypertension, heart failure, and so on there are numerous kinds of heart infections with various sorts of indications. These days there are excesses of computerized procedures to identify coronary illness like information mining, AI, profound learning, and so forth. Thus, in this paper, we will brief prologue to AI methods. In this, we train the datasets utilizing the AI storehouses. The contents of this paper mainly focus on various data mining practices that are valuable in heart disease forecast with the assistance of dissimilar data mining tools that are accessible. If the heart doesn't function properly, this will distress the other parts of the human body such as brain, kidney etc. Heart disease is a kind of disease which effects the functioning of the heart. In today's era heart disease is the primary reason for deaths. WHO-World Health Organization has anticipated that 12 million people die every year because of heart diseases. Some heart diseases are cardiovascular, heart attack, coronary and knock. Knock is a sort of heart disease that occurs due to strengthening, blocking or lessening of blood vessels which drive through the brain or it can also be initiated by high blood pressure.

1.1 Research Background

The major challenge that the Healthcare industry faces now-a-days is superiority of facility. Diagnosing the disease correctly & providing effective treatment to patients will define the quality of service. Poor diagnosis causes disastrous consequences that are not accepted.

Records or data of medical history is very large, but these are from many dissimilar foundations. The interpretations that are done by physicians are essential components of these data. The data in real world might be noisy, incomplete and inconsistent, so data preprocessing will be required in directive to fill the omitted values in the database.

Even if cardiovascular diseases is found as the important source of death in world in ancient years, these have been announced as the most avoidable and manageable diseases. The whole and accurate management of a disease rest on the well-timed judgment of that disease. A correct and methodical tool for recognizing high-risk patients and mining data for timely analysis of heart infection looks a serious want.

Different person body can show different symptoms of heart disease which may vary accordingly. Though, they frequently include back pain, jaw pain, neck pain, stomach disorders, and tininess of breath, chest pain, arms and shoulders pains. There is gigantic exploration proceeding to decide coronary illness hazard factors in various patients, various analysts are utilizing different measurable methodologies and various projects of information mining draws near. Measurable examination has recognized the check of danger factors for heart infections tallying smoking, age, circulatory strain, diabetes, complete cholesterol, and hypertension, coronary illness preparing in the family, stoutness,

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 06, June 2023, pp : 437-447

e-ISSN : 2583-1062

Impact Factor : 5.725

and absence of activity. For the counteraction and medical services of patients who are going to have dependent on coronary illness, it is critical to have consciousness of heart infections.

Scientists utilize a few information mining procedures that are available to help subject matter experts or doctors distinguish coronary illness. Usually utilized methods utilized are choice tree, k-closest, and Naïve Bayes. Other distinctive characterization based procedures utilized are sacking calculation, portion thickness, consecutive insignificant streamlining and neural organizations, straight Kernel self-sorting out guide, and SVM (Support Vector Machine).

AI is an arising field today because of an ascent in the measure of information. AI assists with picking up knowledge from a huge measure of information which is awkward to people and now and then additionally outlandish. The investigation's goal is to organize the conclusion test and see a portion of the solid propensities which add to CVD. Moreover, and in particular, unique AI calculations are contrasted agreeing with various execution measurements. In this postulation, physically arranged information is utilized. Manual grouping is either sound or unfortunate. In light of an AI procedure called characterization, 70% of information are managed or prepared and 30% are tried in this proposal. In this manner, various calculations are looked at according to their forecast results.

Fig .1

1.2 Overview of Machine Learning

Man-made intelligence is a subfield of programming and rapidly flooding topic in the current setting and is depended upon to shoot more in the coming days.

Our existence is flooded with data and data is being made rapidly all around the world.

Man-made intelligence, on other hand, produces strong, repeatable results and gains from the past estimation.

The data used for AI is basic of two sorts named data and unlabeled data.

Coordinated learning is moreover orchestrated into two sorts: Regression and Classification.

1.2.1 Dimension Reduction Technique

The number of input features, variables, or columns present in given dataset is known as dimensionality, and the process to reduce these features is called dimensionality reduction.

A dataset contains a huge number of input features in various cases, which makes the predictive modelling task more complicated**.** Because it is very difficult to visualize or make predictions for the training dataset with a high number of features, for such cases, dimensionality reduction techniques are required to use.

Dimensionality reduction technique can be defined as, " It is way of converting the higher dimensions dataset into lesser dimensions dataset ensuring that it provides similar information". These techniques are widely used in machine learning for obtaining a better fit predictive model while solving the classification and regression problems.

Advantages of this technique, by reducing the dimensions of the features, the space required to store the dataset also get reduced, less computation training time is required for reduced dimensions of features, reduced dimensions of features of the dataset help in visualizing the data quickly, It removes the redundant features (if present) by taking care of multicollinearity.

1.3 Machine learning Algorithms

For the purpose of comparative analysis, The following ML Algorithms are discussed.

K- Nearest Neighbor (KNN) , Random Forest (RF), Support Vector Machine (SVM), AdaBoost, Naïve Bayes and Artificial Neural Network (ANN).

The reason to choose these algorithms is based on their popularity.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 06, June 2023, pp : 437-447

e-ISSN : 2583-1062

Impact Factor : 5.725

K-Nearest Neighbor(KNN), one of the simplest ML algorithms based on Supervised Learning technique.

K-NN algorithms assumes the similarity between the new case into the category that is most similar to the available category.

K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K-NN algorithm.

K-NN algorithm can be used for Regression as well as for classification but mostly used it is used for Classification problems. Random Forest is popular ML algorithms that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It based on concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

RF is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting. SVM (Support Vector Machine Algorithm) SVM is one of the most popular Supervised Learning Algorithms, which used for Classification as well as Regression problems. However, primarily, it uis used for Classification problems in ML. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane.

Naïve Bayes algorithms is supervised learning algorithms, which based on Bayes theorem and used for solving classification problems. It mainly used I text classification that includes a high-dimensional training dataset.

Naïve Bayes Classifiers is one of the simple and most effective classification algorithms which helps in building the fast ML models that can be quick predictions.

It is probabilistic classifier, which means it predicts on the basis of the probability of an object. AdaBoost is an ensemble learning method (also kown as “meta-learning”) which was initially created to increase the efficiency of binary classifiers. AdaBoost uses an iterative approach to learn from mistakes of weak classifiers, and turn them into strong ones.

ANN (Artificial Neural Network)

The term “Artificial Neural Network” is derived from Biological neural networks that develop the structure of a human brain. Similar to the Human Brain that has neurons interconnected to one another, artificial neural networks also have neurons that are interconnected to one another in various layers of the networks. These neurons are known as nodes.

Biological Neural Network Vs Artificial Neural Network

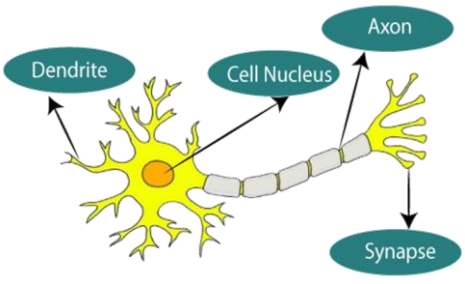

Fig .2

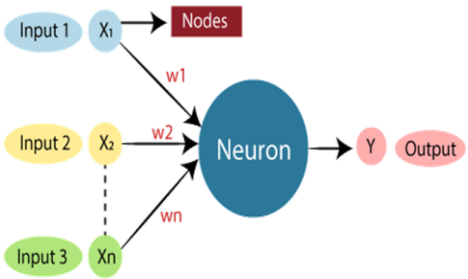

Fig 3

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 06, June 2023, pp : 437-447

e-ISSN : 2583-1062

Impact Factor : 5.725

Convolutional neural network (CNN), It is a type of deep learning algorithm that is particularly well-suited for image recognition and processing task. It is made up of multiple layers, including convolutional layer, pooling layers, and fully connected layers.

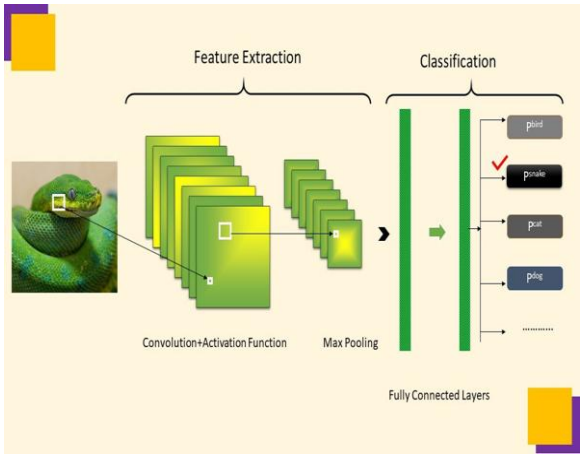

Fig 4

2. PROBLEM IDENTIFICATIONS

Past exploration contemplates have analyzed the utilization of AI methods for the expectation and arrangement of Heart infection. Be that as it may, these investigations center around the specific effects of explicit AI strategies and not on the streamlining of these procedures utilizing upgraded techniques. What's more, not many scientists endeavor to utilize half and half improvement techniques for an enhanced order of AI. The most proposed concentrates in the writing abuse streamlined methods, for example, Particle Swarm Optimization and Ant Colony Optimization with a particular ML strategy, for example, SVM, KNN, or Random Forest. In this work the Fast Correlation-Based Feature Selection (FCBF) technique applied as an initial step (pre-treatment). At the point when all ceaseless ascribes are discretized, the property determination credits applicable to mining, from among all the first ascribes, are chosen. Highlight determination, as a pre-handling step to AI, is successful in diminishing dimensionality, dispensing with insignificant information, expanding learning precision, and improving comprehension of results.

The principal objective of this article is the expectation of coronary illness utilizing diverse grouping calculations, for example, K-Nearest Neighbor, Support Vector Machine, Naïve Bayes, Random Forest, and a Multilayer Perception Artificial Neural Network upgraded by Particle Swarm Optimization (PSO) joined with Ant Colony Optimization (ACO) approaches. The information mining apparatus is utilized to break down information from coronary illness. The principal commitments of this paper are: Extraction of ordered exactness helpful for coronary illness expectation. Remove excess and unessential highlights with Fast Correlation-Based Feature determination (FCBF) strategy. Optimizations with Particle Swarm Optimization PSO then we consider the aftereffect of PSO the underlying estimations of Ant Colony Optimization ACO approach. Comparison of various information mining calculations on the coronary illness dataset. Identification of the best exhibition-based calculation for coronary illness forecast.

3. OBJECTIVES

Heart diseases are currently a major cause of death in the world. This problem is severe in developing countries in Africa and Asia. A heart disease predicted at earlier stages not only helps the patients prevent it, but I can also help the medical practitioners learn the major causes of a heart attack and avoid it before its actual occurrence in patient. In this work, we propose a method named CardioHelp which predicts the probability of the presence of cardiovascular disease in a patient by incorporating a deep learning algorithm called convolutional neural networks (CNN). The proposed method is concerned with temporal data modeling by utilizing CNN for HF prediction at its earliest stage. We prepared the heart disease dataset and compared the results with state-of-the-art methods and achieved good results. Experimental results show that the proposed method outperforms the existing methods in terms of performance evaluation metrics. The achieved accuracy of the proposed method is 97%.

The goal of this investigation is to adequately foresee if the patient experiences coronary illness. The wellbeing proficient enters the info esteems from the patient's wellbeing report. The information is taken care of into model which predicts the likelihood of having coronary illness. In the clinical field, AI can be utilized for finding, location and forecast of different sicknesses.

The primary objective of this work is to give a device to specialists to identify coronary illness as right on time. Predict whether a patient should be diagnosed with heart disease.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 06, June 2023, pp : 437-447

e-ISSN : 2583-1062

Impact Factor : 5.725

Fig 5

4. METHODOLOGY

This section describes the different methods and materials used for this study i.e., research approach, research design and implementation, data collection and tools used for this study.

4.1 Research Approach

This thesis examines the empirical relationship between the set of features such as age, sex and blood pressure with the probability of being diagnosed with heart disease

Therefore, this thesis is a quantitative case study.

Researcher Robert K. Yin defines “a case study as an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used.

Quantitative case study approach is chosen in this thesis because the idea of this study is to investigate available heart disease datasets using a number of statistical methods and running machine learning algorithms on them, so as to find out which one of the machine learning algorithms give better results. Similar studies have been done in the past. The motive of this study is to replicate and extend the previous studies.

4.2 Research Design

This section illustrates the research design; it describes the actual flow of the entire research

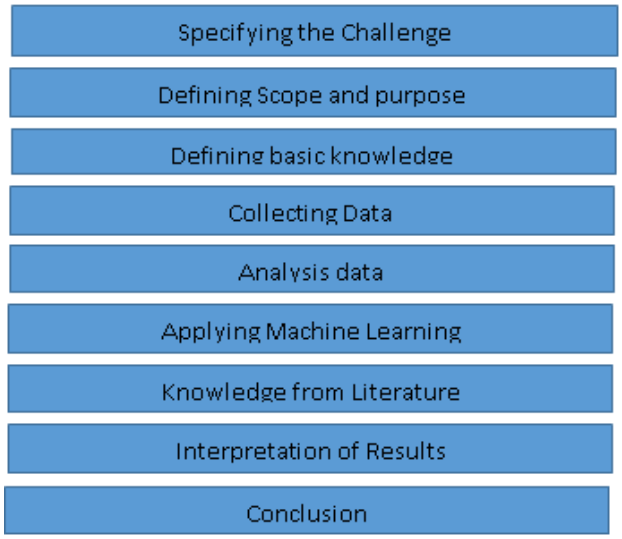

Figure 6. Stepwise flow of the Process

The initial step of this examination was to utilize an accessible coronary illness informational index and contrast distinctive AI calculations with comprehend their exhibition dependent on various execution measurements. In the subsequent advance, the degree of the investigation and its motivation was characterized. The extent of the investigation was characterized according to consider reason, assets, and timetable. The examination's motivation was to comprehend the AI calculations' conduct specifically coronary illness informational collections and to attempt to derive the outcomes. The assets incorporate a PC utilized for the examination, this is critical in characterizing the extent of the investigation in light of the fact that the PC utilized characterizes how a lot and how quick information can be broke down.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 06, June 2023, pp : 437-447

e-ISSN : 2583-1062

Impact Factor : 5.725

4.3 Data Collection and Description

The information is gathered from the UCI AI archive. The informational index is named Heart Disease Data Set and can be found in the UCI AI store. The UCI AI storehouse contains a huge and fluctuated measure of datasets that incorporate datasets from different areas. This information is generally utilized by the AI people group from learners to specialists to comprehend information observationally.

There are actually 76 attributes in the dataset but only 14 attributes are used for this study, these 14 attributes.

Table 1

Features Description	Description
Age	Age Age in years
Sex	Gender instance (0 = Female, 1 = Male)
Cp	Chest pain type (1: typical angina, 2: atypical angina, 3: nonanginal pain, 4: asymptomatic)
Trestbps	Resting blood pressure in mm Hg
Chol	Serum cholesterol in mg/dl
Fbs	Fasting blood sugar > 120 mg/dl (0 = False, 1= True)
Restecg:	Resting ECG results (0: normal, 1: ST-T wave abnormality, 2: LV hypertrophy)
Thalach	Maximum heart rate achieved
Exang	Exercise induced angina (0: No, 1: Yes)
Oldpeak	ST depression induced by exercise relative to rest
Slope	Slope of the peak exercise ST segment (1: up-sloping, 2: flat, 3: down-sloping)
Ca	Number of major vessels colored by fluoroscopy (values 0 - 3)
Thal	Defect types: value 3: normal, 6: fixed defect, 7: irreversible Defect
Num	Diagnosis of heart disease (0: Healthy, 1: Unhealthy)

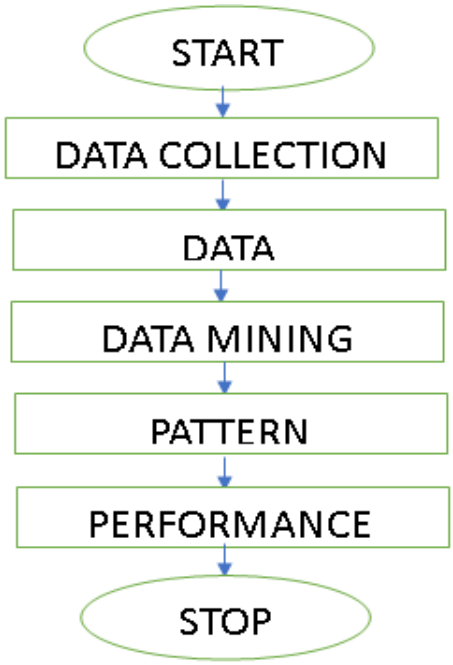

Fig 7

4.4 Performance Metrics

In a machine learning context, performance metrics is the measurement of algorithm on how well algorithm performs based on different criteria such as accuracy, precision, recall etc.

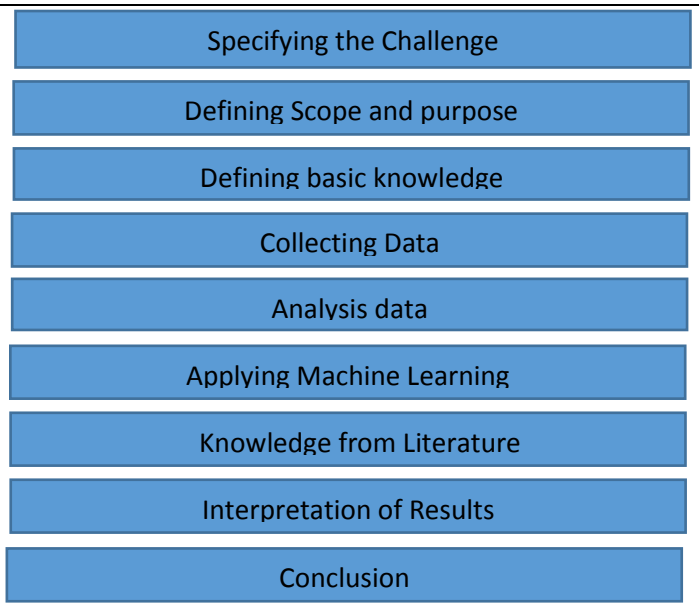

Fig 8

5. RESULT AND DISCUSSION

This section deals with the results obtained from the study of Cleveland dataset and also a comparative analysis of algorithms. After all pre-processing, descriptive and exploratory analysis the data set was employed on different machine learning algorithms, also called models.

5.1 Result Overview- There are two classes found in Scikit-learn machine learning library called Label Encoder and One Hot Encoder. Label Encoder basically transforms the categorical values into numbers which are ordinal in nature. In data set used for this study, there are categorical variables such as Cp, chest pain type which is represented as 1, 2, 3 and 4. 1, 2, 3 and 4 does not have ordinal relationship with each other therefore it gives wrong results when applied directly to machine learning algorithms. Thus, One Hot Encoder is used to encode chest pain type values into binary values, this resolves the issue of ordinality. In this data set the dependent variable or the value to be predicted is multi class. It ranges from 0 to 4. But for this study, multiclass dependent variable is converted into binary class.

5.2 Experiments- Different machine learning models were experimented using Cleveland dataset. In this study initially data set was modelled without feature selection and the results were obtained and in the second phase, data set was modelled only with features obtained from SBS. All experiments included methods like k-fold cross validation and parameter tuning. K-fold cross validation is a technique used in a dataset to avoid over-fitting and under-fitting of the model and parameter tuning is technique which assists to find the best parameters for the model being used.

5.2.1 Training set- Training set is the portion of data in which the model is trained. In this study, 70 percent of data was used for training. In general, in machine learning communities, it is a norm to used 60 to 70 percent of data for training but it varies diversely according to the need and purpose of the experiment. In data training, often the accuracy of training is high, meaning the model. shows high level of accuracy performance in the training set but when tested against the test set, the performance is poor. So to avoid performance error, k-fold cross validation was used. In k-fold cross validation, for example 10-fold cross validation, training set is split in 10 parts and from each 10 part, training and test set is defined and model is employed and the result of all the 10 parts are averaged, this helps to minimize the over fitting and under fitting of the data.

5.2.2 Test set- The test set is the portion of data where the model is tested, it is often the dependent variable of the data. In this study, 30 percent of the data was used for testing. When cross validated data is tested, it will perform better or worse depending on the model used. So, to ensure every model is functioning in its optimum, a technique named parameter tuning was used. Scikit-learn library contains the class called Grid Search CV which performs the parameter tuning.

5.2.3 Results with various Parameters

The aim of the entire project was to test which algorithm classifies diseases the best. This section includes all the results obtained from the study and introduces the best performer according to various performance metrics.

First, performance was obtained by using 10-fold cross validation in training set. Secondly, performance was obtained just by using model without any parameter tuning, third parameters were tuned and fourth model was calibrated. The following tables shows the results.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 06, June 2023, pp : 437-447

e-ISSN : 2583-1062

Impact Factor : 5.725

Table 2 : KNN gave best results

Algorithms Used	Accuracy	Precision	Recall	F1-score	ROC AUC	Log loss
KNN	0.84	0.89	0.73	0.83	0.90	0.42
SVM	0.82	0.81	0.79	0.82	0.89	0.41
Random Forest	0.67	0.64	0.63	0.66	0.85	1.30
Naïve Bayes	0.80	0.81	0.72	0.79	0.87	1.51
CNN	0.85	0.90	0.80	0.86	0.92	1.62

Referring Table , CNN gave best results in the 10-fold cross validation followed

Following Table, shows the performance of algorithms in test set without parameter tuning.

Table 3 . Evaluation of algorithms in test set without parameter tuning.

Algorithms Used	Accuracy	Precision	Recall	F1-score	ROC AUC	Log loss
KNN	0.82	0.83	0.81	0.81	0.81	0.44
SVM	0.79	0.78	0.78	0.78	0.78	0.45
Random Forest	0.69	0.68	0.68	0.68	0.69	1.0
Naïve Bayes	0.80	0.81	0.72	0.79	0.87	1.51
CNN	0.85	0.82	0.82	0.83	0.82	1.55

Referring to Table, CNN is the best performer followed. In the test set, Naïve Bayes performed better than in the training set but log loss of Naïve Bayes is high than SVM. It can mean that accuracy of Naïve Bayes classifier is not entirely true, classifier is over fitted.

Table, shows the performance of algorithms in test set using parameter tuning.

Table 4, Evaluation of algorithms in test set using parameter tuning.

Algorithms Used	Accuracy	Precision	Recall	F1-score	ROC AUC	Log loss
KNN	0.78	0.83	0.78	0.77	0.77	0.54
SVM	0.80	0.82	0.80	0.80	0.79	0.43
Random Forest	0.73	0.74	0.74	0.73	0.73	0.49.
Naïve Bayes	Nil	Nil	Nil	Nil	Nil	Nil

In Table, Naïve Bayes' row is Nil because it does not need parameter tuning, there are no parameters to be tuned in Naïve Bayes. After parameter tuning, KNN performance goes low. SVM, followed by Naïve Bayes and Random Forest are best performer after parameter tuning. In figure, performance of log loss is excluded because it is negative in nature and can exceed one in its value, which is not supported by other performance metrics.

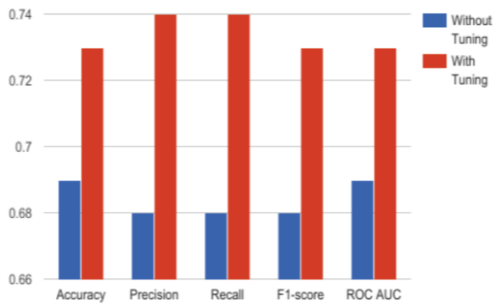

Figure 9. Random Forest with and without parameter tuning

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 06, June 2023, pp : 437-447

e-ISSN : 2583-1062

Impact Factor : 5.725

In figure, Blue bars represents performance without tuning and red bars represents after tuning. It is seen that after parameter tuning random forest performance is much better than before, still its performance is low in this dataset. Now all of these algorithms are calibrated using Platt's scaling or also known as sigmoid function. Table, shows the results after the calibration.

Table 5, Evaluation of algorithms in test set, parameter tuned using calibration

Algorithms Used	Accuracy	Precision	Recall	F1-score	ROC AUC	Log loss
KNN	0.78	0.78	0.78	0.78	0.77	0.46
SVM	0.79	0.80	0.79	0.79	0.78	0.43
Random Forest	0.75	0.73	0.73	0.72	0.72	0.5.
Naïve Bayes	0.80	0.81	0.80	0.80	0.79	0.48
CNN	0.85	0.88	0.75	0.87	0.88	1.00

In Table, after calibration CNN outperforms all the algorithms and also log loss of Naïve Bayes has decreased from 1.10 to 0.48.

Table, shows the performance of algorithms after they were tuned and calibrated and feature selected using SBS. This helps to compare the model performance before and after feature selection.

Table 6 . Evaluation of tuned and calibrated algorithms with feature selection

Algorithms Used	Accuracy	Precision	Recall	F1-score	ROC AUC	Log loss
KNN	0.78	0.78	0.78	0.78	0.77	0.44
SVM	0.85	0.86	0.86	0.86	0.85	0.37
Random Forest	0.78	0.78	0.78	0.78	0.77	0.45.
Naïve Bayes	0.79	0.80	0.79	0.79	0.78	0.92
CNN	0.85	0.87	0.79	0.87	0.87	1.01

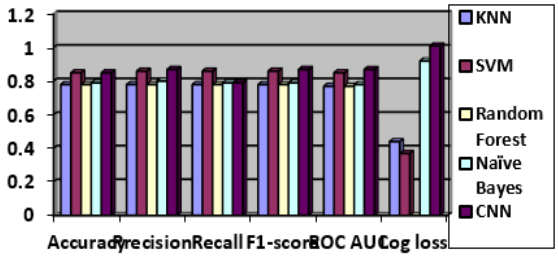

Figure 10

Referring to Table, CNN is the best classifier followed by Naïve Bayes, KNN and Random Forest.

Table 7 , Evaluation of Artificial Neural Network with parameters

Algorithms used	Accuracy	Precision	Recall	F1-score	ROC AUC	Log loss
ANN	0.91	0.92	0.89	0.90	0.87	0.23
CNN	0.92	0.94	0.90	0.93	0.88	0.60

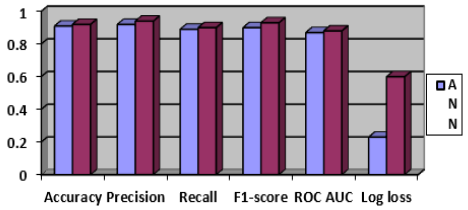

Figure 11. Comparison of CNN and ANN.

ANN and CNN in Table, uses 3 hidden layers and also uses 'adam' algorithm as an optimizer.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 03, Issue 06, June 2023, pp : 437-447

e-ISSN : 2583-1062

Impact Factor : 5.725

5.3 Comparison Results

From all the tables above, different algorithms performed better depending upon the situation whether cross validation, grid search, calibration and feature selection is used or not. Every algorithm has its intrinsic capacity to outperform other algorithm depending upon the situation. CNN perform there best in terms of accuracy, Precision ,Recall For example, Random Forest performs much better with a large number of datasets than when data is small while Support Vector Machine performs better with a smaller number of data sets. Performance of algorithms decreased after boosting in the data which was not feature selected while algorithms were performing better without boosting in not feature selected data. The performance of algorithms increased after boosting in the data which was feature selected while algorithm performance decreased after boosting in feature selected data. This shows the necessity that the data should be feature selected before applying boosting. For the comparison of dataset, performance metrics after feature selection, parameter tuning and calibration is used because this is standard process of evaluating algorithms. In Table 5.6, performance metrics of classifiers are added except log loss because lower the log loss better is the classifier, so log loss is subtracted from the added performance metrics and then averaged. Average value of SVM, Random Forest and Naïve Bayes are 0.63, 0.59 and 0.63 respectively. This shows SVM and Naïve Bayes are performing on average.

Table 8

Algorithm	**Accuracy**	Precision	Recall	F1 Score
Logistic Regression	85.25	72.11	71.23	68.22
Naive Bayes	85.25	71.12	70.13	69.15
Linear SVM	81.97	75.36	73.99	72.11
KNN	67.21	65.32	64.21	67.66
Decision Tree	81.97	74.32	73.21	72,56
XGBoost	85.25	76.66	75.12	77.86
Neural Network	80.33	72.12	72.56	74.12
Random Forest	95.08	81.12	80.31	79.26
Proposed CNN	95.99	81.14	81.23	80

6. CONCLUSIONS AND FUTURE SCOPE

The objective of the undertaking was to contrast the calculations and diverse execution measurements utilizing AI. All information was pre-handled and utilized for the expectation on the test. Each calculation performed better in certain circumstances and more regrettable in another. CNN is the conceivable models to work best in the dataset utilized in this investigation. To have the option to analyze coronary illness precisely utilizing AI has numerous significances. Various gadgets can be fabricated which will screen the heart-related exercises and analyze the illness. These gadgets will end up being useful where coronary illness specialists are not accessible. With additional examination, AI can likewise be utilized to analyze coronary illness before human specialists can do. One of the principal accomplishments of this task was that undertaking assisted with understanding the calculations better. At the point when tried through different circumstances, the calculations performed diversely which helped me to comprehend the calculation's working system. This postulation can be the primary learning step in coronary illness determination with AI and it very well may be broadened further for future examination. There are a few impediments to this investigation basically the information base of the creator, furthermore, the apparatuses utilized in this examination, for example, the handling intensity of the PC, and thirdly the time restriction accessible for the examination. This sort of study requires the condition of-workmanship assets and skill in particular areas.

7. REFERENCES

[1] Senthilkumar Mohan, Chandrasegar Thirumalai, Gautam Srivastava Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques‖, Digital Object Identifier 10.1109/ACCESS.2019.2923707, IEEE Access, VOLUME 7, 2019 S.P. Bingulac, On the Compatibility of Adaptive Controllers,‖ Proc. Fourth Ann. Allerton Conf. Circuits and Systems Theory, pp. 8-16, 1994. (Conference proceedings)

[2] Sonam Nikhar, A.M. Karandikar" Prediction of Heart Disease Using Machine Learning Algorithms" International Journal of Advanced Engineering, Management and Science (IJAEMS) Infogain Publication,[Vol-2, Issue-6, June- 2016].I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[3] Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Prof. Kailas Devadkar (PhD)," Prediction of Heart Disease Using Machine Learning", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018).IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1- 5386-0965-1

[4] Abhay Kishore1, Ajay Kumar2, Karan Singh3, Maninder Punia4, Yogita Hambir5," Heart Attack Prediction Using Deep Learning", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 04 | Apr-2018.

[5] A.Lakshmanarao, Y.Swathi, P.Sri Sai Sundareswar," Machine Learning Techniques For Heart Disease Prediction", International Journal Of Scientific & Technology Research Volume 8, Issue 11, November 2019.

[6] Mr.Santhana Krishnan.J, Dr.Geetha.S," Prediction of Heart Disease Using Machine Learning Algorithms",2019 1st International Conference on Innovations in Information and Communication Technology(ICIICT),doi:10.1109/ICIICT1.2019.8741465.

[7] Avinash Golande, Pavan Kumar T," Heart Disease Prediction Using Effective Machine Learning Techniques", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019.

[8] V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja," Heart disease prediction using machine learning techniques: a survey", International Journal of Engineering & Technology, 7 (2.8) (2018) 684-687.

[9] V. Manikantan and S. Latha, "Predicting the analysis of heart disease symptoms using medicinal data mining methods", International Journal of Advanced Computer Theory and Engineering, vol. 2, pp.46-51, 2013.

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

[11] S. M. S. Shah, S. Batool, I. Khan, M. U. Ashraf, S. H. Abbas, and S. A. Hussain, ''Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis,'' Phys. A, Stat. Mech. Appl., vol. 482, pp. 796–807,2017. doi:10.1016/j.physa.2017.04.113.

[12] Stephen F. Weng, Jenna Reps, Joe Kai1, Jonathan M. Garibaldi, Nadeem Qureshi,―Can machine-learning improve cardiovascular risk prediction using routine clinical data?‖, PLOS ONE | https://doi.org/10.1371/journal.pone. 0174944 April 4, 2017.

[13] N. Al-milli, Back propagation neural network for prediction of heart disease, ''J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp.131–135, 2013.

[14] S. Abdullah and R. R. Rajalaxmi, ''A data mining model for predicting the coronary heart disease using random forest classifier,'' in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22–25.

[15] V. Krishnaiah, G. Narasimha, N. Subhash Chandra, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review" IJCA 2016.

[16] K.Sudhakar, Dr. M. Manimekalai "Study of Heart Disease Prediction using Data Mining", IJARCSSE 2016.

[17] NagannaChetty, Kunwar Singh Vaisla, NagammaPatil, "An Improved Method for Disease Prediction using Fuzzy Approach", ACCE 2015.

[18] Vikas Chaurasia, Saurabh Pal, "Early Prediction of Heart disease using Data mining Techniques", Caribbean journal of Science and Technology,2013