ISSN No:-2456-2165

Coronary Illness Hazard Prediction using Machine Learning (CIHPML)

Y V NageshMeesala*, N Sai Sankar[2], M Abhishek[3] and P Rahul[4]
Raghu Engineering College, Visakhapatnam, India

Abstract:- Cardio vascular diseases are one of the major causes of death globally.It is very important tofind a precise and well founded approach to automate the accomplished work and thus carrying out effective management. Many researchers used several data mining techniques to understand and help in diagnose heart disease. In order to decrease the deaths from heart disease, you must have a fast and precise detection technique. Early prediction can help people change their lifestyle. It also ensures proper medical treatment if needed. In order to drop down the death rate of heart diseases, a rapid and precise techniquesare needed. The proposed work predicts the possibilities of heart diseases by implementing various number of data mining techniques such as logistic regression, K nearest, decision trees, support vector machine. A web based system is developed in this paper, that can determine whether a person is likely to get affected with heart disease or not based his health factors. It is found that the Support Vector Machine achieved a maximum accuracy of 86.76% against other implemented ML algorithms.

Keywords:- Machine Learning, Health care, Cardio Vascular Diseases, Modelling and Training, Cardiologist.

I. INTRODUCTION

The main objective of our paper is prediction of heart condition within the next 10 years by exploitation totally different data processing tools. It is of course the most important organ in our body. It supplies blood to all the parts in our body. As per the World Health Organization, Cardio Vascular Disease (CVD) is the prime explanation for mortality worldwide.

It is further analyzed that 23.3 million people would be deceased by the year 2030 due to CVDs globally. Many researchers have done significant studies and identified the CVD causing factors like hereditary, stress levels, living standards, dietary habits, physiological factors and so on [1].

Generally in the United States, in every 40 seconds a person dies with heart attack. This clearly indicates the need to concentrate more on heart related issues. Consequently, this led to an expenditure of over $200 billion per year in the United States alone.This also going to increase at a high rate in the upcoming years as well.

To minimize the risk rate of people dying for CVD there is a need for automated technique that can assist in CVD research and examining the attributes and symptoms of CVD that leads to early prediction and prevention of the diseases. The advent and development in the digital technology in medical diagnosis has increased the dimensions to detect the pre-existence of the disease. Over the last few decades there have been lot of efforts to train machines to replicate the human brain which gave rise to the concept of machine learning [2].

***Corresponding author Y V Nagesh Meesala:** Raghu Engineering College, India E-mail: nagesh.myv@raghuenggcollege.in

Unhealthy diet, excess alcohol consumption also has a greater risk on functioning of our heart. Early detection of heart diseases are often challenging and difficult to determine in most of the cases. Thus, the increase in the computer-aided detection helped to analyze and understanding the diagnoses in the medical field as well.

Today the major challenge is the efficient and accurate prediction of these diseases. One capable was is the Machine Learning that is training and testing with the help of python and its libraries. Data mining is the process that can bring outusefulinformation from enormous amount of raw data collected.

So, this MLis nothing but the processing of data and handling large datasets efficiently. Thus, in the medical field it can help a lot in the prevention and detection of the diseases.

The AI and ML algorithms have shown their potential ability to determine the existence of the CVD by accurate data mining on the health records and medical history of families of the patients.

Since the predictions are related to CVD there are higher chances of risk and possibilities for the patient to end up with a serious condition and may also have a heart failure. This happens due to the fact that medication is not prescribed in the beginning itself and the next visit of the patient is not known as one may feel that he is safe and may take longer time for the next diagnosis.

This ML is a part of AI, that can become proficient by itself thus capable of making decisions and any predictions. Initially, we train the algorithm with the required dataset as the input, then the model learns and capable of determining patterns from the dataset. Afterwards, we can send new data, so that it can predict to which class data belongs to.

ISSN No:-2456-2165

II. LITERATURE SURVEY

A. Machine Learning techniques, models in CVD prediction

J. Mishra et.al [3] has used machine learning methods for forecasting the chronic diseases in the patients and then used Adam as an optimizer and has given a review on the growing popularity of AI and ML algorithms in field of medical sciences especially in prediction of the diseases.

Himanshust al.[4] briefly discussed about large and small data set of heart diseases prediction. They shared that small data set take minimum time for training as well as testing and performed prediction using SVM and KNN algorithm. It also discussed about prediction of heart diseases and prove that some algorithms of machine learning does not perform better for accurateness predication.

M. Tarawneh et al. [5] has urged the need for an expert system to be served as analysis tool for discovering the information and patterns in heart diseases. They proposed a prediction system for heart disease by combining techniques that involved data mining in an algorithm for prediction and claimed to give more accurate results.

F. Miao et al [6] built up a thorough danger model utilizing improved arbitrary endurance woodland (iRSF) used to foresee mortality of cardiovascular breakdown. In view of investigation made with clinical data set of 8059 patients, 32 danger factors, including meds, socioeconomics, and clinical, lab data and has built up the danger model. A few cases are made which support that the trail consequences of the danger model created are utilized as significant instrument by specialists for forecast of mortalitiy of cardiovascular breakdown.

Stephen F. Wenget.al, [7]implementedvarious ML techniques topredict theCVD diseases.It shows that various ML algorithms are useful for increasing the exactness in identificationof CVD diseases, but it needs more patient records. So the more the data, the better the results.

Marjia Sultana et.al, [8] traversed various datasets for Heart disease illness and determined the usage of various Machine Learning algorithms with them. Obviously, datasets are to be preprocessed before applying any Machine Learning algorithms. They also suggest the various features that plays important for accuracy determination.

PrajaktaGhadge et al. [9] conducted an explorationon heart attack prediction. It uses big data and various modelling techniques as well. Therefore, the system can identify any hidden knowledge and pattern in the vast data.

AH Chen et al. [10] developed the modelwhich can help doctors predict status of the heart disease with the help of clinical data. It used various programming languages. Programming languages like C and C# are used and with an accuracy of 80%.

B. Web based CVD

The designed framework usually collects data from the user in a simple manner. It collects information about the patients health condition through signs like circulatory strain, blood glucose levels etc..that are often varying. [11].

The following are some of the devices through which a user can enter his details.

- **Device 1:** A mobile phone that has an active internet connection, can be used to enter the details of the patient i.e, Age, Gender, Blood Pressure, BPM etc.. A user can go through the website and there's no need to login or register to it. He can just enter the details that are needed and click on submit to get the results. A user can go back and reset by clicking the Home button.
- **Device 2:** A user can also enter his details from a laptop/desktop with internet connection. The process is same as in mobile phone and laptop as well. So a user can enter the corresponding details by visiting the website. Without any registration, a patient can easily enter his details and check his status for the next ten years.

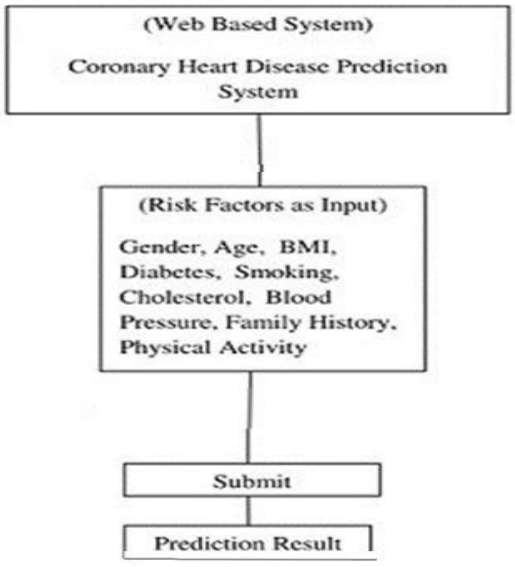

Fig. 1: Block Diagramof Proposed System

ISSN No:-2456-2165

C. Web based Implementation:

Coronary Illness hazard identification can be done when the patient enters the details of his medical record i.e, Age, Gender, Blood Pressure, BPM etc. One can enter his details from any device with active internet connection.[11].

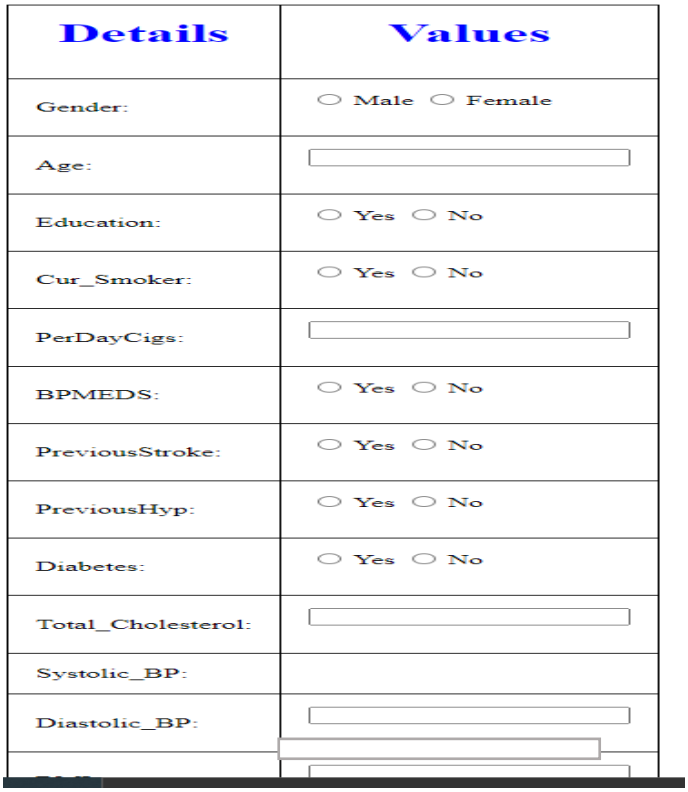

Fig. 2: Input given to system

So after entering his details any patient can check his status and can determine whether he needs doctor assistance or not.

Fig. 3: Result

D. Aim:

To predict with the outcome of whether a patient ought to be determined to have a coronary illness in the next ten years.

E. Attributes

Gender: male or female(Nominal)
Age: Patient's Age
Education: Is he/she currently pursuing education
Cur_Smoker: If the patient is a current smoker or not
PerDayCigs: cigarettes smoked per day
BPMEDS: blood pressure medication of the patient
PreviousStroke: patient previousstroke information
PreviousHyp: If patient was hypertensive or not
Diabetes: whether the patient had diabetes
Total_Cholesterol: total cholesterol level
Systolic_BP: blood pressure
Diastolic_BP: blood pressure
BMI: Body Mass Index
Heart_Rate: Heart Rate of the patient
Glucose: glucose level of the patient

ISSN No:-2456-2165

F. Target variable to predict:

Risk of coronary heart disease(10 years) – {1:'yes', 0:'no'}

Here, Table 1 shows the rows and columns of the dataset

	Gender	Age	Education	CUr_Smoker	Per Day Cigs	BP MEDS	PreviousStroke	Previous Hyp	Diabetes	totChol	Systolic_BP	Diastolic_BP	BMI	Heart_Rate	GLUCOSE	TENYEARCHD
0	1	39	1	0	0	0	0	0	0	195	106	70	26.97	80	77	0
1	0	46	0	0	0	0	0	0	0	250	121	81	28.73	95	76	0
2	1	48	0	1	20	0	0	0	0	245	127.5	80	25.34	75	70	0
3	0	61	1	1	30	0	0	1	0	225	150	95	28.58	65	103	1
4	0	46	1	1	23	0	0	0	0	285	130	84	23.1	85	85	0

Table 1: Rows and Columns of the dataset

- **Information Analysis:** Correlation is nothing but an indication about the changes between two variables.The relationship of factors utilizing positive Correlation Matrix in Figure 2. Using this we can see there is a positive connection between diabetes and glucose levels in the body. This looks good because the more is the glucose in the body leading to diabetes.

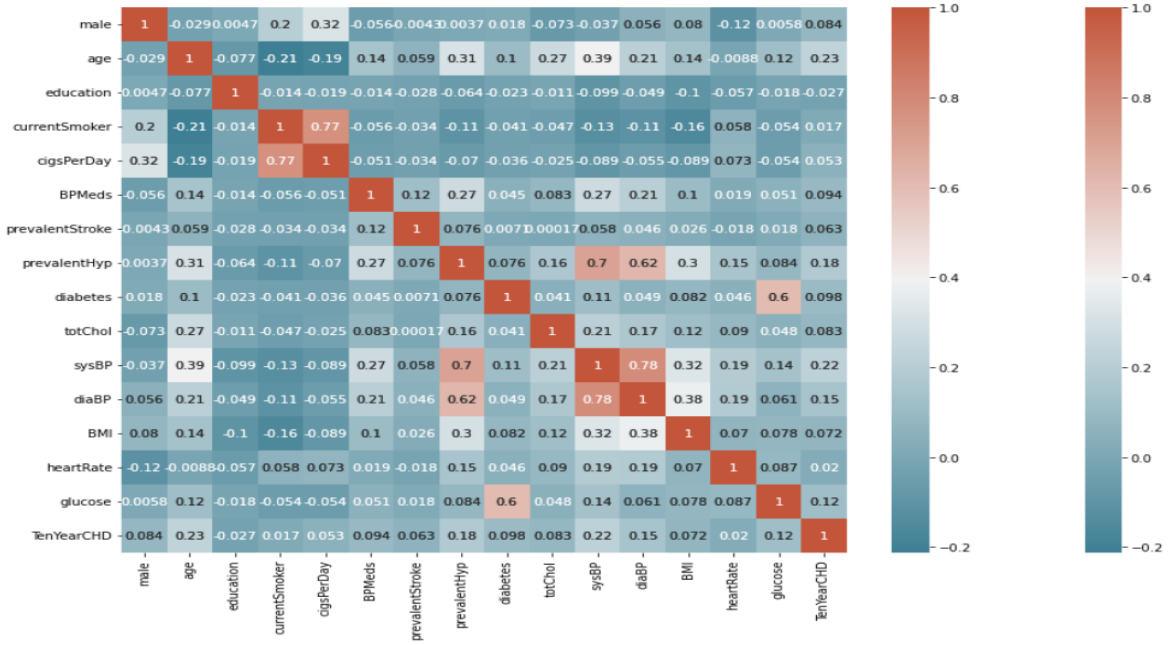

Fig. 4: Information is emphatically or adversely associated with our indicator(target).

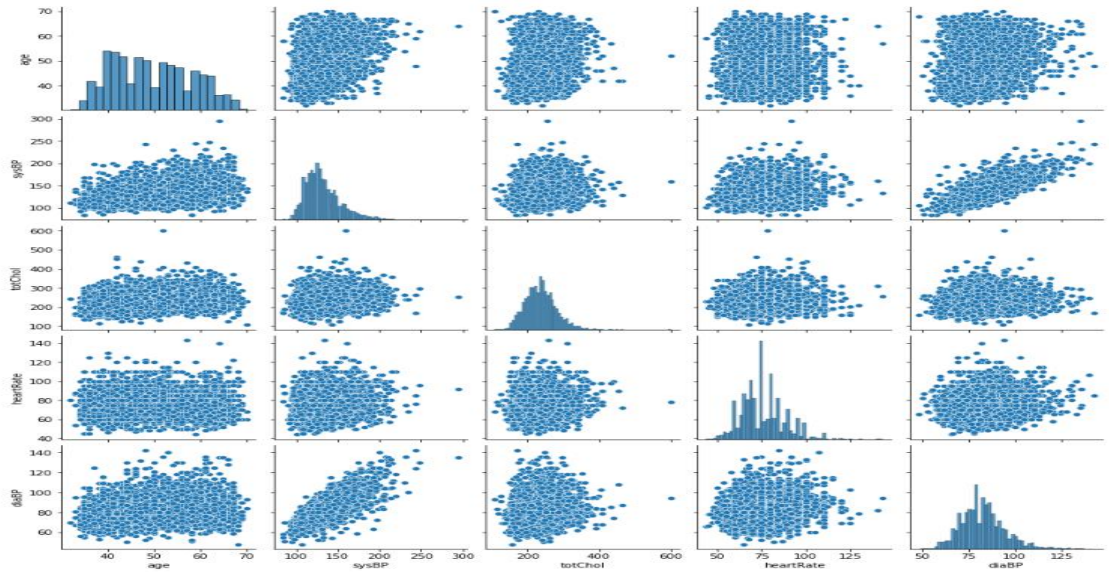

Fig. 5: Plot we make a more modest pair plot with just the consistent factors, to jump further into the connections Pairplots are likewise an extraordinary method to quickly see the relationships between all factors

ISSN No:-2456-2165

- **Modelling and training:** Demonstrating different algorithms on the aimed information which gives the most important elevated precision. We will analyze the precision of comparative on different machine learning algorithms.

➢ **Model 1: Logistic Regression**

The target variable is selected and the probability of the variable occurrence is predicted by implementing this algorithm. This is mostly used for binary problems. The data coded of the target variable is in binary nature (0 or 1) [2]. Here, table 2 shows the result.

	Precision	Recall	F1-score	Support
0	0.7	0.64	0.67	714
1	0.58	0.64	0.61	548
accuracy			0.64	1262
macro avg	0.64	0.64	0.64	1262
weighted avg	0.65	0.64	0.64	1262

Table 2: Accuracy with Logistic Regression.

➢ **Model 2: Support Vector Machine**

They are amazingly adaptable managed AI calculations which are utilized both for characterization and relapse. [2].Here, table 3 shows the report of Support Vector Machine.

	Precision	Recall	F1-score	Support
0	0.93	0.82	0.87	714
1	0.79	0.92	0.85	548
accuracy			0.86	1262
macro avg	0.86	0.87	0.86	1262
weighted avg	0.87	0.86	0.86	1262

Table 3: Accuracy with SVM

➢ **Model 3: K-NN (K-Nearest Neighbors)**

The classification techniques of this algorithm based on considering the likeliness of the target variable with other corresponding variables to get the impact of their coordination. The performance of this supervised learning algorithm lies in the data size, if the data size is large then the performance is low and vice versa. [13].

	Precision	Recall	F1-score	Support
0	0.94	0.73	0.82	714
1	0.73	0.94	0.82	548
accuracy			0.82	1262
macro avg	0.84	0.84	0.82	1262
weighted avg	0.85	0.82	0.82	1262

Table 4: Accuracy with K-NN

➢ **Model 4: Decision Trees**

They use the tree like structures in the model to decide on the outcomes which contain control statements. The impact of the decision will also be calculated with the probability of estimating the result.These structures are really helpful for predicting the outcomes that one is expecting. They also help to understand and determine the results in a much better way[2].

	Precision	Recall	F1-score	Support
0	0.8	0.72	0.76	714
1	0.68	0.77	0.72	548
accuracy			0.74	1262
macro avg	0.74	0.74	0.74	1262
weighted avg	0.75	0.74	0.74	1262

Table 5: Accuracy with Decision Tree

G. Highlight Importance

This section gives the importance factor which is causing the disease with the score comparison of the selected variables. The accuracy of the model mostly depends on the selection of the factors and the ranges of the values which a variable assigned with. The calculated result of the each factor leads to selection and inclusion of the factor in the model if the result is low then factor is omitted if it is showing desirable results it is included. The correlated factors of the included factor are examined to consider them for including them in the decision.

III. RESULT

Score how accommodating each element was in our model

Highlight: 0, Score: 0.01959
Highlight: 1, Score: 0.12968
Highlight: 2, Score: 0.01703
Highlight: 3, Score: 0.01205
Highlight: 4, Score: 0.05312
Highlight: 5, Score: 0.00861
Highlight: 6, Score: 0.00343
Highlight: 7, Score: 0.01830
Highlight: 8, Score: 0.00930
Highlight: 9, Score: 0.12227
Highlight: 10, Score: 0.14504
Highlight: 11, Score: 0.11663
Highlight: 12, Score: 0.12935
Highlight: 13, Score: 0.09710
Highlight: 14, Score: 0.11848

IV. CONCLUSION

The heart related disease detection, prediction and medication aspects are always challenging and crucial for patient. It is always risk if the heart patients are not treated in time and if they are far from their home town or if the regular doctored is not available. The Support vector machine was the best performing model in terms of accuracy and the F1 score. Its high AUC shows that it has a high true positive rate and the SMOTE technique helped in improving the models sensitivity by balancing the dataset, this is when compared to the performance metrics of other models on different notebooks on the same dataset. Furthermore, the advancements and availability of internet enables users to access irrespective of their economical and social factors.

This model not only increases the confidence of the patient but also enhances the doctor's ability to treat with utmost care. The treatment module of the model is completely based on the quality and availability of internet connection, thus care should be taken to maintain the internet available.

REFERENCES

[1.] Baban.U. Rindhe , Nikita Ahire , RupaliPatil, ShwetaGagare , Manisha Darade , Heart Disease Prediction Using Machine Learning, International Journal of Advanced Research in Science, Communication and Technology, Volume 5, Issue 1, May 2021, ISSN (Online) 2581-9429

[2.] B Padmajaa , ChintalaSrinidhib , KothaSindhuc , KalaliVanajad , N M Deepikae , E Krishna RaoPatrof, Early and Accurate Prediction of Heart Disease Using Machine Learning Model, Turkish Journal of Computer and Mathematics Education, Vol.12 No.6 (2021), 4516-4528

[3.] J. Mishra and S. Tarar, Chronic Disease Prediction Using Deep learning, Singapore: Springer 2020, pp. 201-211.

[4.] Himanshu Sharma and M A Rizvi. (2017). Prediction of Heart Disease using Machine Learning Algorithms: A Survey, International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 8 , IJRITCC August 2017.

[5.] M. Tarawneh and O. Embarak, Hybrid approach for heart disease prediction using data mining techniques, Acta Sci. Nutritional Health, vol. 3, no. 7, pp. 147-151, Jul. 2019.

[6.] F. Miao, Y -P, Cai, Y -X, Zhang, X -M Fan and Y Li, Predictive modelling of hospital mortality for patients with heart failure by using an improved random survival forest, IEEE Access, vol. 6, pp. 7244-7253, 2018.

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

[8.] M. Sultana, A. Haider, and M. S. Uddin, Analysis of data mining techniques for heart disease prediction, 2016 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol. ICEEICT 2016, 2017

[9.] Ghadge, P., Girme, V., Kokane, K. and Deshmukh, P. (2016) Intelligent Heart Attack Prediction System Using Big Data. International Journal of Recent Research in Mathematics Computer Science and Information Technology, 2, 73-77.

[10.] AH Chen, SY Huang, PS Hong, CH Cheng, and EJ Lin,2011, HDPS: Heart Disease Prediction System, Computing in Cardiology, ISSN: 0276-6574, pp.557-560

[11.] Harish Rajora, Narinder Singh Punn, Sanjay Kumar Sonbhadra , and SonaliAgarwal, Web based disease prediction and recommender system, Indian Institute of Information Technology Allahabad, India.

[12.] ElliackinFigueiredo , Mariana Macedo And Hugo Valadares Siqueira Swarm Intelligence For Clustering A Systematic Review With New Perspectives On Data Mining, Https://Doi.Org/10.1016/J.Engappai.2019.04.007, Received 10 August 2018; Elsevier Ltd.

[13.] Abualigah, L.M., Khader, A.T., Al-Betar, M.A., Alomari, O.A., 2017a. Text Feature Selection With A Robust Weight Scheme And Dynamic Dimension Reduction To TextDocument Clustering. Expert Syst. Appl. 84, 24–36