**HEART ATTACK PREDICTION USING MACHINE LEARNING**

Annaji Kuthe^{*1}, Achal M. Chavan² and Neha S. Lambat²

^{1,2,3}Department of Computer Technology, KDKCE, Nagpur, India.

Article Received on 03/01/2022

Article Revised on 24/01/2022

Article Accepted on 13/02/2022

***Corresponding Author**

Annaji Kuthe

Department of Computer
Technology, KDKCE,
Nagpur, India.

ABSTRACT

Within the current era, Heart Failure (HF) is one of the common diseases that can lead to dangerous situation. Because of this disease almost 26 million of patients are affecting. As per the heart consultant and surgeon's point of view, it is difficult to predict the heart failure on

right time. Classification and predicting models are there, which aid the medical field and can illustrate how to use the medical data efficiently. Since diagnosis could be a sophisticated task and plays a significant role in saving human lives thus it has to be done accurately and with efficiency however there's lack of effective tools to find hidden relationships and trends in e-health information. This paper aims to improve the Heart Failure prediction accuracy using UCI (Universal Chess Interface) heart disease dataset. We use, various machine learning techniques such as to understand the data and predict the Heart Failure differences in medical database. An appropriate and accurate computer-based automated decision support system is required to reduce the cost for achieving clinical tests. As well the results and relative study showed that, the current work better the previous accuracy score in predicting heart disease. The integration of the machine learning model presented in this study with medical data methodology is beneficial to predict the Heart Failure.

KEYWORDS: Machine Learning, Logistic regression, Heart disease, Random forest, K-NN, Decision Tree.

INTRODUCION

Daily increasing development in information technology caused in vital growth in sciences. One of the sciences is life science. Victimization computer science techniques altogether subjects of this branch of science particularly vas diseases created it attainable to style health

professional systems. By taking attention to extend in new diseases and conjointly extension of technologies, the identification of diseases gone on the far side the interior treatment vogue, and also the most efforts of doctors and specialists is concentrated on early prediction of diseases victimisation out there signs.^[7]

Heart diseases are the most reasons for death worldwide. Consistent with the survey of the globe Health Organization (WHO), 17.5 million total world deaths occur due to heart attacks and strokes. Over seventy-fifth percent of deaths from vas diseases occur principally in middle-income and low-income countries. Also, eightieth percent of the deaths that occur because of CVDs are due to stroke and attack.^[1]

Therefore, detection of internal organ abnormalities at the first stage and tools for the prediction of heart diseases will save a great deal of life and facilitate doctors to style an efficient treatment set up that ultimately reduces the morbidity because of vas diseases. Because of the event of advance attention systems, scores of patient information are these days on the market (i.e. massive information in Electronic Health Record System) which might be used for planning prognostic models for vas diseases. Data processing or machine learning may be a discovery methodology for Associate in Nursing analysis massive data from a various perspective and encapsulating it into helpful information. “Data Mining may be a non-trivial extraction of implicit, antecedent unknown and probably helpful info regarding data”.^[2]

Nowadays, an enormous quantity of knowledge concerning illness diagnosing, patients etc. are generated by attention industries. Data processing provides variety of techniques that discover hidden patterns or similarities from information.^[3] Therefore, during this paper, a machine learning formula is planned for the implementation of a heart condition prediction system that was valid on 2 open access heart condition prediction datasets. Higher cognitive process is another widespread space. The importance of higher cognitive process has been according to several researchers in varied fields. A number of them being E-Learning, foretelling, websites style, development.^[4, 5, 6 and 7] WSNs additionally notice place in varied fields that integrates higher cognitive process and are with success applied in utility based mostly, multi criteria, fusion metric and context enabled higher cognitive process etc.

LITARATURE REVIEW

Diagnosis of heart diseases is a significant and boring task and also an important duty in

medical science, which requires extreme attention. However there is some tools for data extraction and analysis. Also existence of huge set of medical data leads to correct diagnosis of disease. Using medical data including age, sex, blood pressure, and blood sugar, it is possible to increase the possibility of heart diseases prediction. These data must be collected in organized manner, which could be used for integrating the prevention system [8]. In different countries using artificial intelligence techniques and various algorithms, predicting this type of death -heart disease- is somewhat possible. In Iran many efforts in this subject made by cooperation of software and medical communities. So the current study focuses on field studies by aim of decreasing the cost and early prediction of events happened for heart patients in Iran. One of important methods in this field is clustering. In clustering, the data splits to some clusters, in such a way that the data in every cluster have maximum similarity with each other and minimum similarity with data of other clusters. So using clustering data will show that every cluster that has the patient, could help us in predicting that if he/she is under heart attack risk or not.

Using this technique and going to additional precise designation for heart patients is our aim. We all know that ancient clump strategies like K-Means typically judges on information victimization the gap between them. However during this study the best objection is victimization this property, as a result of out there information concerning heart patients [9] includes binary and nominal information. Therefore victimization varied improved clump strategies, we are able to use alternative metrics [10] rather than specializing in distance between information, to target qualitative properties, to extend the exactness and gain additional correct designation. By examination and clump the techniques and algorithms that utilized in heart diseases field and its designation, In this research paper using different algorithmic rules we tends to find out that ,which algorithm will gives the higher accuracy score to predict the condition of Heart Failure.

DATA SET INFORMATION

The name of the dataset is heart.csv. There are unit 303 instances during this dataset, wherever the cases are unit either individuals having heart condition or they're healthy. Among 303, 165 (54.45%) cases are unit individuals with heart condition and 138 (45.54%) are unit individuals while not heart condition. The quantity of attributes is fourteen, each

row represents one patient record. The first 13 attributes are used in the prediction part and the last column its output in the binary form 0 and 1. Features include age, sex, chest pain, trtbps, thol, fbs, restecg, thalach, exang, oldpeak, slp, caa, thall, Output. In the figure (Figure.1: Dataset) below shown the dataset we used.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1

Figure 1: Dataset.

METHODOLOGY

A. System Architecture

For the proposed study dataset was taken from Kaggle site. Then it was download in excel file using comma separated format. Data has processed by python programming using Jupyter notebook. The dataset contains 14 clinical features. Different types of python libraries such as pandas, Sklearn, NumPy, matplotlib are used for processing the algorithms. Using explorative data analysis technique data was analysed in jupyter notebook. We are using 4 different Machine Learning Techniques. We develop the app using flask to show the result where the doctor or patient fill their medical report data as input and it will predict the result weather the patient will suffer from heart Failure of not. In the Figure below (Figure.2:Architecture) general process of data has been shown.

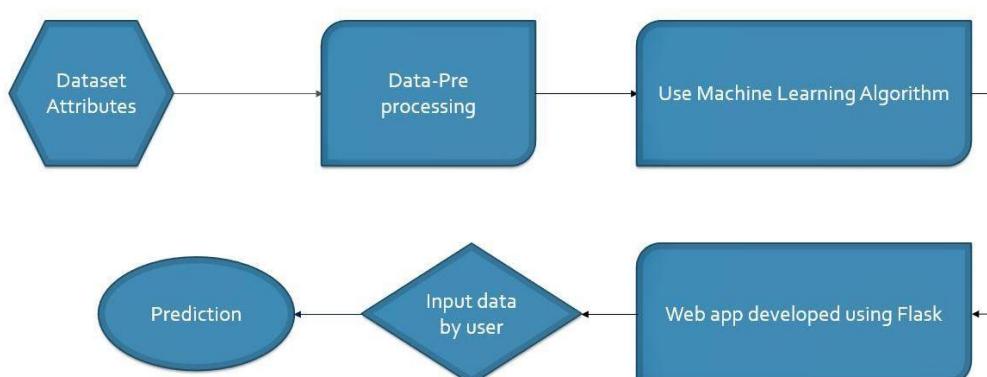


Figure 2: Architecture.

1. Logistic Regression

Logistic regression is applied mathematics analysis technique wont to predict an information price supported previous observations of a dataset. Supplying regression has become a crucial tool within the discipline of machine Learning. The approach permits associate in nursing formula getting used in an exceedingly machine learning application to classify incoming information supported historical information. As additional relevant information comes in, the formula ought to improve at predicting classifications inside information sets. Supplying regression may play a task in information preparation activities by permitting information sets to be place into specifically predefined buckets throughout the extract, transform, load (ETL) method so as to stage the data for analysis.

2. Decision Tree

In this algorithm at a specific node, a split of knowledge happens. Thus the most effective attribute to separate is to be known. Once the split for every price, a child node is formed. For every child node if the set is pure then it might stop otherwise a algorithmic split happens. This algorithm is prime down and goes on in dividend conquer manner. There are three varieties of nodes in an exceedingly call tree specifically root node, branch node and leaf node. Leaf node represents a category. Partitioning of knowledge is stopped once the subsequent conditions are satisfied.

- i. Once all samples for a given node belong to identical category.
- ii. No attributes ought to be remained for. Further partitioning and so as to classify leaf andopt for majority balloting is employed.
- iii. No information samples ought to be left. Once the choice tree is made mistreatment that the testing set, it's foretold by traversing across the tree every, for every sample and selectingthe suitable price at each node.

3. Random Forest

The combination of learning models increases the classification accuracy. This technique is called bagging and the main idea of this is to average noisy and unbiased models in order to create another model with a lower variance in terms of classification. This algorithm works as a large collection of de-correlated decision trees and is based on the above mentioned bagging technique. From the sample set a lot of subsets with random values are created. Thus using the subsets a corresponding decision tree is created for each subset. Random forests correct for decision trees way of overfitting to their training set.

4. K-NN

K-nearest neighbors could be a straightforward algorithmic program that stores all accessible cases and classifies new cases supported a similarity measure k-nearest neighbors algorithmic program (k-NN) could be a non-parametric technique used for classification and regression. In each cases, the input consists of the k-nearest coaching examples within the feature area. The output depends on whether k-NN is employed for classification or regression: In k-NN classification, the output could be a category membership. An object is assessed by a plurality vote of its neighbors, with the item being appointed to the category commonest among its k nearest neighbors (k could be a positive whole number, usually small). If k = 1, then the object is solely appointed to the category of that single nearest neighbor. In k-NN regression, the output is that the property price for the item. This price is that the average of the values of k nearest neighbors. K-NN could be a variety of instance-based learning, or lazy learning, wherever operate is simply approximated regionally and every one computation is deferred till classification. The k-NN algorithmic program is among the only of all machine learning algorithms. Both for classification and regression, a helpful technique may be accustomed assign weight to the contributions of the neighbors, so that the nearer neighbors contribute additional to the common than the additional distant ones. As an example, a typical weight theme consists in giving every neighbor a weight of $1/d$, wherever d is that the distance to the neighbor.

B. Web Application

Basically Flask is used for developing a small size website. Flask is easy to make API using python. It is a set of functions and predefined classes used to connect with the system software and handle inputs and outputs. It simplifies the life of a developer while giving them the ability to use certain extensions and makes the online applications scalable and maintainable. As of now, we have develop a model i.e model.pkl which can predict result using the dataset based on a various attribute of the data.

RESULT

Analysis of data shows that the performance of Logistic Regression Machine is more accurate than other Machine learning Algorithms (Figure.3:Accuracy-score) for the specified testing data set and sample training data set used in this study. This study was based on the idea that values used for various parameters affecting heart disease. Logistic Regression model which is having a low error rate has proven to be useful in Hearth disease prediction among other

prediction techniques. In which the KNeighbor has the 68.42% of accuracy-score which is less and found to be error in prediction. Decision Tree has the accuracy of 75% which is way better than K-Neighbor and the Random forest has the accuracy-score of 78.94% which is as near as result of Logistic Regression which has accuracy-score of 86.84%.



Figure 3: Accuracy-score.

CONCLUSION

This heart condition prediction model with an accuracy of 78.94% can facilitate individuals , particularly medical professionals to scale completely different eventualities. Patients on the opposite hand also can consult a doctor beforehand and bear check-up and so can stop the incidence of any heart condition. During this paper, we tend to projected 4 ways during which analysis was done and results were achieved. The conclusion that we tend to found is that machine learning performed higher analysis. For the 13 features which were in the dataset, Linear Regression performed better in the ML approach when data pre-processing is applied. The Computing time was additionally reduced that is helpful, for deploying a model.

It was additionally distinguished that the dataset ought to be normalized. Otherwise, the training model gets overfitted generally and therefore the accuracy achieved isn't sufficient once a model is evaluated for real-world knowledge issues which may vary drastically to the dataset on that the model was trained. It had been distinguished that the applied mathematics analysis is additionally necessary once a dataset is analysed and it ought to have a normal distribution, so the outlier's detection is additionally necessary and a method referred to as Isolation Forest is employed for handling this. The formula applied by North American nation in ANN architecture accumulated the accuracy that we have a tendency to compare with the

various researchers. The dataset size will be accumulated so deep learning with varied alternative optimizations will be used and brighter results will be achieved. Machine learning and varied alternative improvement techniques can also be used in order that the analysis results can once more be accumulated. Additional other ways of normalizing the information will be used and therefore the results will be compared. And additional ways in which could be we have a tendency toll be found wherever we might integrate heart-disease-trained ML models with bound transmission for the convenience of patients and doctors.

REFERENCE

1. M.Gjoreski, A.Gradišek, M. Gams, M. Simjanoska, A. Peterlin, and G. Poglajen, - Chronic heart failure detection from heart sounds using a stack of machine-learning classifiers, in Proceedings 2017 13th International Conference on Intelligent Environments, IE 2017, 2017, pp. 1419.
2. G. Savarese and L. Lund, —Global Public Health Burden of Heart Failure,|| Card. Fail. Rev., vol. 3 no 1, 2017.
3. E. J. Benjamin, P. Muntner, and et al. Alonso, Alvaro, —Heart Disease and Stroke Statistics— 2019 Update: A Report From the American Heart Association,|| Circulation, vol. 139, no. 10, 2019.
4. Annaji M. Kuthe, Rohini Choudhari, Pratiksha Dehanikar et al. E-Learning Management WebApp with Video Conferencing. International Research journal of Engineering & Technology (IRJET) 2021, Vol 08, Issue 06, pp. 1935-1941.
5. A. M. Kuthe, Sujata S. Wasnik, Priti G. Kumbhare et al. Prevention of Suicide Risk and Predicting Suicidal Behaviors by machine learning. Wutan Hutan Jisuan Jishu 2021, Vol XVII, Issue-I, pp. 563-567.
6. A. M. Kuthe, Mukta Dhimole, Prachi bante et al. An Anti-Theft Android Device safeguard Application. International Journal of Scientific Research in Engineering and Management (IJREM) may 2021, Vol 05, I-05, pp.1-5.
7. Samiksha Lade, Disha kalambe, Ritika Jisnani, Annaji M. Kuthe. Road Mishap (Accident Detector). International Research journal of Engineering & Technology (IRJET) Sep 2019, Vol 06, Issue 09, pp. 521-523.
8. M. Ramaraj and T. A. Selvadoss, —A Comparative Study of CN2 Rule and SVM Algorithm and Prediction of Heart Disease Datasets Using Clustering Algorithms,|| Netw. Complex Syst., vol. 3, no. 10, pp. 1–6, 2013.

9. A. Gavhane, G. Kokkula, I. Pandya, and P. K. Devadkar, —Prediction of Heart Disease Using Machine Learning,|| in Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018, 2018, pp. 1275–1278.
10. H. Murthy and M. Meenakshi, — Dimensionality reduction using neuro-genetic approach for early prediction of coronary heart disease, in International Conference.