Predicting of Heart Disease using Machine Learning

Ardra Sreekumar[1]
[1]Student,
Dept. Of Computer Science & Engineering,
Mangalam College of Engineering, India,

Celina Joythi[2]
Student,
Dept. of Computer Science & Engineering,
Mangalam College of Engineering, India,

Ms. Roja Thomas
Assistant Professor,
Dept. of Computer Science &
Engineering, Mangalam College of Engineering, India,

***Abstract*- Cardiovascular disease are very difficult to diagnose in the modern health field so it is important to understand the human heart health. In this paper, wepropose efficient genetic algorithm hybrid with the back propagation technique approach for heart disease prediction. Heart disease is one of the major aliments that people suffer from today. It is difficult to detect with a simple observation. Treatments with early monitoring can also cause death in patients. Computer based and supporting system can be developed tomake as good and effective treatment modes. In the usual way many hospitals information patient chart sheet my be contained likes images ,test chart, and numbers but it is difficult to diagnose all the conditionsof a patient through this. The main objective of this paper is to develop a prototype which can determine and extract unknown knowledge related with heart disease from a past heart disease database record. It can help to reduce the costs of treatment.**

Keywords— Genetic algorithm, back propagation technic

1. INTRODUCTION

Coronary illness is one of the most well-known reasons of death in India or other Asian nations. In 2003 approx 17.3 million individuals kicked the bucket all around the globe and out of this, 10 millionwere simply because of the coronary illness. Along without changing way of life there are many such factors, for example, smoking, liquor, corpulence, hypertension, diabetes and so forth. which are liable for the gamble of having a heart issue. In any case, with the disdain studies, with the presentation of man-made reasoning and clinical sciences, we can in fact help in forestalling any such sort of sicknesses. For making a great choice, AI helps in extricating significant information from tremendous data sets which are accessible in clinics. There are numerous sort of grouping strategies like K-closest neighbor, choice trees like CART, C4.5, CHAID, J48, ID3algorithm and so on yet every one of these are powerless classifiers which need the assistance ofsacking and supporting procedures to get to the next level their exhibitions .In this paper, different sorts ofstrategies which have been applied in the forecast of heart sicknesses or characterization has been examined and a proposed procedure of half breed procedure has been given which can be executed in future to have a precision of practically 100 percent or withleast blunder. The proposed framework will beexecuted in MATLAB R201 .Coronary illness isperhaps the most well-known reasons of death inIndium or other Asian nations. In 2003 approx. 17.3 million individuals passed on all around the globe also, out of this, 10 million were exclusively becauseof coronary illness. There are many changing way oflife there are many such factors like smoking, liquor,heftiness, hypertension, diabetics and so forth whichare depend able for the gamble of having a heart issue. Computerized reasoning and clinical sciences, it canreally help in extricating applicable . There is a heft ofsecret data in this information that isn't yet investigated which lead to an significant question ofhow to make valuable data out of the information. Sothere is need of making an astounding undertakingwhich will help experts anticipate the coronary illness before it happens. The fundamental goal of this paper is to foster a model which can decide and separate obscureinformation (examples and relations) related with coronary illness from a past coronary illness information base record

II LITERATURE SURVEY

With the rising passing counts because of coronary illness , a framework is expected set up for precisely anticipating coronary illness . The investigation has likewise been introduced on the fundamental of innovation utilized, The discoveries of the review recommend that the proposed diagnostics framework can be doctors to anticipate coronary illness precisely.By staying away from the model from overfitting and under fitting , of can show great execution under fittingit can show great execution on both the datasets that is preparing information and testing information. Anyway , the greater part of these methods center around the preprocessing of highlights just . We center around both that is refinement of elements and disposal of the issues presented by the prescient modelthat is the issues of under fitting and overfitting . The expectation of endurance of coronary illness has been a difficult examination issue for clinical society . In view of the information , we worker three famous information

mining calculations to foster the expectation models utilizing the cases . The outcomesdemonstrated that the SVM is the best indicator with precision brain network emerged to be the second withexactness and the choice trees models . The expectation examination is the procedure of information mining which can be anticipate further potential outcomes situated in the ongoing data .

A Analysis of coronary disease and prediction of heart attack in coal mining regions using data mining techniques.

Coronary illness (HD) is a significant reason for horribleness and mortality in the cutting edge society. Clinical conclusion is critical yet convoluted task that ought to be performed precisely and effectively. This review breaks down the Behavioral Risk Factor Surveillance System, study to test whether self- revealed cardiovascular illness rates are higher in Sindarin coal mining locales in Andhra Pradesh state,India, contrasted with different districts after control for different dangers. Subordinate factors incorporate self-revealed proportions of being determined to have cardiovascular sickness (CVD) or with a particular type of CVD including (1) chest torment (2) stroke and

(3) respiratory failure. Heart care concentrate on indicates 15 credits to anticipate the dreariness. Next to ordinary ascribes other general credits BMI (Body Mass Index), doctor supply, age, nationality, instruction, pay, and others are utilized forexpectation.

B .Analysis of various heart disease prediction techniques.

The forecast examination is the strategy of information mining which can foresee further conceivable outcomes situated in the ongoing data. This exploration work, depends on coronary illness forecast in information mining. The dataset expectation informational index has 13 number of qualities for thecoronary illness forecast. In the past exploration work,the SVM classifier is applied for the coronary illness expectation. Because of huge number of traits in the dataset, SVM classifier can't order every one of the properties because of which precision is low for the expectation.

C. Combination data mining methods with new medical data to predicting outcome of Prediction of cardiovascular disease .

Forecast of cardiovascular sickness is a basic test in the space of clinical information examination. AI (ML) has been demonstrated to be compelling in helping with simply deciding and expectations from the enormous amount of information created by the medical care industry. We have additionally seen ML methods being utilized in ongoing improvements in various region of the Internet of Things (IoT). Different investigations give just a brief look into foreseeing coronary illness with ML procedures. In this paper, we propose an original strategy that targetstracking down critical elements by applying AI procedures bringing about working on the precision inthe forecast of cardiovascular illness

III. EXISITING MODEL

Heart disease is one of the most common reasons of death in India or other Asian countries. In 2003 approx. 17.3 million people died all over the globe andout of this, 10 million were only due to coronary heartdisease. There are many changing lifestyle there are many such factors such as smoking, alcohol, obesity, high blood pressure, diabetics etc. which are responsible for the risk of having a heart problem. Artificial intelligence and medical sciences, it can actually help in extracting relevant . There is a bulk ofhidden information in this data that is not yet exploredwhich give rise to an important query of how to make useful information out of the data. So there is necessityof creating an excellent project which will help practitioners predict the heart disease before it occurs.The main objective of this paper is to develop a prototype which can determine and extract unknown knowledge (patterns and relations) related with heart disease from a past heart disease database record

1) DATA MINING TECHNIC

Data mining techniques aid in the extraction and analysis of complex medical data utilising a variety of methods. These are also beihmmng used by medical practitioners. Other fields, such as cancer detection, use minute approaches.as well as the stroke A variety of techniques have been used by researchers. Artificial Neural Network, BP, and other methods of machine learning Genetic method for the (Back-Propagation Algorithm)the objective of optimization Back-Propagation has been employed in neural networks byone of the systems. The most effective predictionalgorithm is a network If the data and the model have a nonlinear connection desired result BP algorithm hasthe following characteristics Its tolerant of noisy data and other outliers and adapts to theming the medical records, there is a reference to it.Following are the steps: The data that has been normalised

2) DECISION TREE

There are many decision tree algorithms and among them the most popular is J48 which uses pruningtechnique to build a good decision tree. Pruning is a method which tries to eliminate the over fitting data which is not so relevant in making a decision and leadsto poor prediction. At last, a tree is built to provide flexibility and accuracy balance.

3) NAÏVE BAYES

This classifier utilizes restrictive autonomy which expresses that a trait XYSTEEM on a given class is not subject to the upside of different properties are variable as its depends on base hypothesis every one of the three calculations where carry out of informational index compressing of 303 records fore preparing and 270 for testing reason in preprocessing the most one be recognized apparatus of weka for example supplant. Missing of channel is utilize

IV. METHODOLOGY

A. Proposed System

In this paper the proposed new technique to coronary illness. Cardio vascular sicknessare extremely challenging to analyze in the model wellbeing field so it is essential to comprehend the human heart wellbeing therapies with early observing can likewise cause demise in patients. We propose a proficient genera spasm calculations crossover with the back engendering. Illness expectation utilizing patient treatment history and wellbeing information by applying information mining and AI strategies is continuous battle for as far back as many years. Manyworks have been applied information mining methodsto neurotic information or clinical profiles for expectation of explicit infections. These methodologies attempted to foresee the reoccurrence of infection. In present day frameworks from what wegive patients all that we can find out what their confidence is, that is the new framework we are making. This is depending the patient's age, blood bunch, and so forth by postulations things the framework was originating before the kind of cardiovascular infection and which sort of medicines they need.

DATASET FEATURES

(i) Age—age of patient in years, sex—(1 male; 0 female).

(ii) Cp—chest pain type.

(iii) Trestbps—resting blood pressure (in mm Hg onadmission to the hospital). The normal range is120/80 (if you have a normal blood pressure reading, it is fine, but if it is a little higher than it should be, youshould try to lower it. Make healthy changes to your lifestyle).

(iv) Chol—serum cholesterol shows the amount oftriglycerides present. Triglycerides are another lipidthat can be measured in the blood. It should be less than 170 mg/dL (may differ in different Labs).

(v) Fbs—fasting blood sugar larger than 120 mg/dl (1 true). Less than 100 mg/dL (5.6 mmol/L) is normal, and 100 to 125 mg/dL (5.6 to 6.9 mmol/L) is considered prediabetes.

(vi) Restecg—resting electrocardiographic results.

(vii) Thalach—maximum heart rate achieved. The maximum heart rate is 220 minus your age.

(viii) Exang—exercise-induced angina (1 yes). Angina isa type of chest pain caused by reduced bloodflow to the heart. Angina is a symptom of coronary artery disease.

(ix) Oldpeak—ST depression induced by exercise relative to rest.

(x) Slope—the slope of the peak exercise ST segment.

(xi) Ca—number of major vessels (0–3) colored by fluoroscopy.

(xii) Thal—no explanation provided, but probably thalassemia (3 normal; 6 fixed defects; 7 reversible defects).

(xiii) Target (T)—no disease 0 and disease 1, (angiographic disease status).

B.ALGORITHM

k-closest neighbors(KNN) calculation is as sort of directed ml calculations which can be utilized for both order as well as replace persistence issue,. Be that as itmay it is primarily utilize for grouping persistence issues in industry

C. Modules

Admin
- o Login
- o View approve/reject doctors
- o Train dataset
- o Test data with dataset
- o View user/patient details
- o View complaints and send reply
- o View top rated doctors

Doctors
- o Register
- o Login if approved
- o Manage fee
- o Manage consulting time
- o View booking

View patient details
Accept/reject booking
View payments
- o Chat with patient/users
- o View rating and reviews

Users
- o Register
- o Login
- o Input data to test heart disease prediction with machine learning
- o View prediction history
- o View doctors Sort

by top rated
View nearby doctors
View fee amount

- o Book a doctor

View booking status
Make payment
- o Chat with doctors
- o Send complaints and view reply
- o Review and rate a doctor

Tables login

- o login_id
- o username
- o password
- o user_typedoctors
- o doctor_id
- o login_id
- o first_name
- o last_name
- o house_name
- o place
- o landmark
- o qualification
- o phone
- o email

o status
users
o user_id
o login_is
o first_name
o last_name
o house_name
o place
o phon
o email
complaints
o complaint_id
o user_id
o complaint
o reply
o date_timehistory
o history_id
o user_id
o result
o date_timefee
o fee_id
o doctor_id
o amount
o date_time consulting_times
o consulting_id
o doctor_id
o day
o start_time
o end_time
o date_time
bookings
o booking_id
o user_id
o consulting_id
o date_time

o book_date
o status
payments
o payment_id
o booking_id
o date_time
o status
chat
o chat_id
o sender_id
o sender_type
o receiver_id
o receiver_type
o message
o date_time
ratings
o rating_id
o user_id
o doctor_id
o rate
o review
o date_time

D. System Architecture

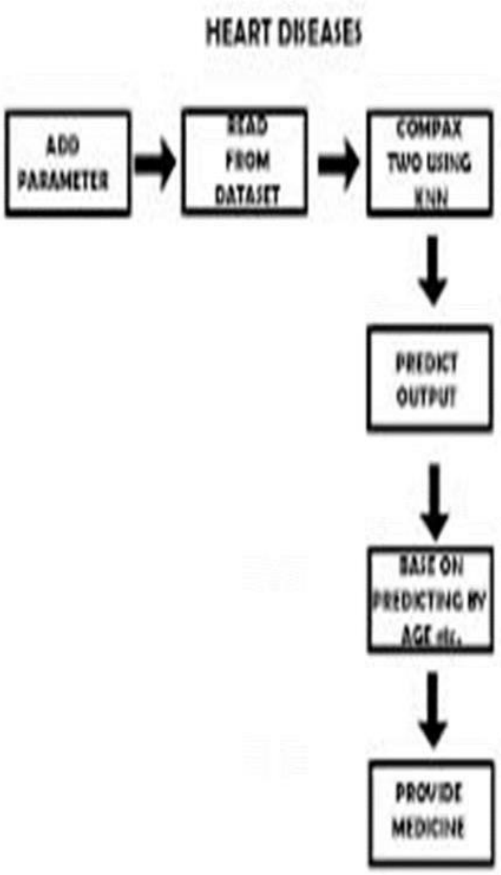

Fig 1

Determination of heart disease type and the complication analyse through the parameter input.That parameter totally converted into a group of dataset . by the algoritham MNN read the dataset & also analyse the complication and its number or its problems. Through this details algoritham can be predict the exact output.

Based on the prediction analyse the patients situation a, age , pasr treatment type . by this information can beprovide the perfect medicine

V. FUTURE SCOPE

Coronary illness is term used to portray a star groupingof condition that can influence the heart and its valves, vessels, structure electrical framework or coronary corridors condition that for inside in the extents of coronary illness incorporate cardiovascular arrhythmias hypertension cardiovascular breakdown coronary supply route valve jumble and conjunctional heart surrenders among others through these sickness impact the heart diversely a definitive issue with all assortments of coronary illness is that somehow , they can upset the imperative siphoning activity of the heart

VI. CONCLUSION

The conclusion which we found is that machine learning algorithms performed more better in this analyzation many researches are previously suggestedthat we should use ML ware the data set is not that large. Heart disease

describes a range of conditions that effect the patients heart. Now a days cardio vascular disease are the leading cause of death. Various unhealthy activities are the reason for the increase in the risk of heart disease like high cholesterol obesity increase in triglycerides levels hypertensions etc. There are open sources for accessing the patients records and researches various computer technologies will be used for doing the correct diagnosis of the patients and detect this diseases to stop it from becoming fatal.

ACKNOWLEDGEMENT

The authors wish to thank Principal Dr.Vinodh PVijayan, Dr.Ranju S Kartha, H.O.D, Computer Science Department,for the proper guidance, valuable support, and helpful comments during the proofreading.

REFERENCES

[1] Agrawal R, Srikant R. Fast algorithms for mining association rules. In: Proceedings of 20th international conference verylarge data bases, VLDB. Vol. 1215, pp. 487–499; 1994.

[2] Akbaş KE, Kivrak M, Arslan AK, Çolak C.Assessment of association rules based on certainty factor: an application on heart dataset, in 2019 Inter- national artifcial intelligence and data processingsymposium (IDAP) (pp. 1–5). IEEE; 2019.

[3] Altaf W, Shahbaz M, Guergachi A. Applications ofassociation rule mining in health informatics: a survey. Artif Intell Rev. 2017;47(3):313–40.

[4] Alwidian J, Hammo BH, Obeid N. WCBA: weighted classifcation based on association rules algorithm for breast cancer disease.Appl Soft Comput. 2018;62:536–49.

[5] American Heart Association. Heart disease and stroke statistics 2017 at- a-glance. Geraadpleegd van: https://healthmetrics.heart.org/wp-content/ ploads/2017/06/Heart-Disease-and-Stroke- Statistics-2017-ucm_491265.pdf.

[6] Amin MS. Identifying signifcant features and datamining techniques in predicting cardiovascular disease; 2018.

[7] Amin MS, Chiam YK, Varathan KD Identifcation of signifcant fea-tures and data mining techniques in predicting heartdisease. Telem Inform. 2019;36;82–93.

[8] Bashir, S., Khan, Z. S., Khan, F. H., Anjum, A., &Bashir, K. (2019). Improv- ing heart disease prediction using feature selection approaches. In 2019 16th International Bhurban Conference on AppliedSciences and Technology (IBCAST) (pp. 619–623). IEEE.

[9] Cengiz AB, Birant KU, Birant D. Analysis of pre-weighted and post-weighted association rule mining, in 2019 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1–5). IEEE.

[10] Chauhan A, Jain A, Sharma P, Deep V. Heartdisease prediction using evo- lutionary rule learning, in 2018 4th International conference on computa- tional intelligence & communication technology(CICT) (pp. 1–4). IEEE; 2018.

[11] Dey L, Mukhopadhyay A. Biclustering-basedassociation rule mining approach for predicting cancer-associated protein interactions. IET Syst Biol. 2019;13(5):234–42.

[12] Dua, D., Graf, C. UCI machine learning repository [http://archive.ics.uci.edu/ ml]. Irvine, CA: University of California, School of Information and Computer Science; 2019.

[13] Domadiya N, Rao UP. Privacy-preserving association rule mining for horizontally partitioned healthcare data: a case studyon the heart diseases. Sādhanā. 2018;43(8):1–9.

[14] Domadiya N, Rao UP. Privacy preserving distributed association rule mining approach on vertically partitioned healthcare data.Procedia Comput Sci. 2019;148:303–12.

[15] Fitriyani NL, Syafrudin M, Alfan G, Rhee J. HDPM: an efective heart disease prediction model for a clinical decision supportsystem. IEEE Access. 2020;8:133034–50.

[16] Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier; 2011.

[17] Ibrahim SP, Sivabalakrishnan M. An enhanced weighted associative clas- sifcation algorithm without preassigned weight basedon ranking hubs. Int J Adv Comput Sci Appl. 10(10); 2019.

[18] Ibrahim SS, Sivabalakrishnan M. An evolutionary memetic weighted associa- tive classifcation algorithm for heart diseaseprediction. In Recent Advances on Memetic Algorithms and its Applications in ImageProcessing (pp. 183–199). Springer, Singapore; 2020.

[19] James SL, et al. Global, regional, and nationalincidence, prevalence, and yearslived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for theGlobal Burden of Disease Study 2017. Lancet392 (10159), 1789–1858; 2018.