RESEARCH ARTICLE OPEN ACCESS

Analysis of the Occurrence of Heart Disease Using Data Mining Techniques

Pujitha Chowdary, Boddapati
Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur - India

ABSTRACT
Forecast of the event of heart infections in clinical focuses is huge to distinguish if the individual has heart infection or not. Information mining is utilized to recover covered up data in clinical priorities that help to foresee unique disorders. Coronary illness is one of the most well-known sicknesses that lead to death in this world. Every year, 17.5 million individuals are biting the dust because of cardiovascular disease, as per the World Health Association reports. One of the most well-known clinical focuses is that everything specialists don't have equivalent information; what's more, treating their patients gives their own choice that may provide helpless outcomes and lead the patients to death. To conquer such issues, the expectation of heart sicknesses utilizing information mining strategies and AI calculations assumes essential functions for programmed analysis of infection in medical services places. Some machine calculations used to anticipate heart sicknesses are Support Vector Machine, Decision Tree, Naïve Bayes, K-Nearest Neighbor, and Fake Neural Network.

Keywords :- Naïve Bayes, Support Vector Machine, Decision Trees, K-Fold Cross-Validation, Heart Disease, Machine Learning.

I. INTRODUCTION

Information mining is the method of finding significant examples; furthermore, information from an immense measure of information in the information base, information distribution centers, web or information put away in various data stores. Numerous studies must find concealed data from enormous datasets that help partners comprehend and recover their information inside a brief period. Diverse information mining methods are utilized to arrange, anticipate, and bunch information to make right or precise dynamic in numerous associations. In clinical focuses (medical clinics or other clinical priorities), utilizing information mining procedures helps patients distinguish if the person has sicknesses or not for the early programmed determination of patients from their conditions inside the outcome recover in a short time.

Coronary illness is one of the most pervasive disorders that can lead to death, incapacity, and other monetary emergencies in patients who experience coronary illness's ill effects. Every year 17.5 million individuals pass on because of CAD (Coronary Artery Disease) in this world, as indicated by the World Health Organization (WHO) reports. Because computerized advances are quickly developing, medical care communities store enormous information in their information base that is exceptionally perplexing and testing to examinations. Information mining procedures and Machine learning calculations assume crucial functions in various medical care data (clinics and other clinical) focuses.

A forecast is a proper procedure in medical services places where clinicians don't have more information and ability. There are no trained professionals; for example, such clinicians may choose to give a helpless outcome and lead the patients to death. The expectation of coronary illness is utilized for a programmed finding of the infection and gives administrations' adequate characteristics in medical care communities to spare people's lives. The expectation method helps make a clear choice for the partners, especially for experts, to treat patients' sensible choice.

II. HEART DISEASE

The heart is one of the body parts fundamental for each part of the body by coursing or siphoning blood to each body part like a mind. If the heart quits flowing blood to mind diverse neuron nerve framework are kick the bucket, that causes all nerves or tissue goes to other body parts quit working, and passing may happen. The goal of life, at last, relies upon the heart. An appropriately working spirit encourages people to have a good life.

Forecast of cardiovascular sickness is testing and the more convoluted errand to accomplish a programmed analysis of ailment. A considerable measure of information is put away in medical care places that are exceptionally perplexing and testing to investigations. Whether it is a troublesome errand utilizing forecast of heart sicknesses in clinical focuses assume huge parts to spare people's way of life and make dynamic and exact dynamic for partners.

Probably the most well-known heart infections are recorded in the table beneath with their portrayal

TABLE - 1: TYPES OF HEART DISEASE

Types of heart diseases	Description
Arrhythmia	Atypical heart rhythm
Acute coronary syndrome	Blood supply to the heart muscle is swiftly obstructed
Coronary artery disease	Occurred when Arteries supplying blood becomes obstructed
Angina	Chest pain due to a deficiency of blood to the heart muscle
Congenital heart disease	Heart disfigurements that are present at birth
Cardiomyopathy	Heart muscle disease
Rheumatic heart diseases	Rheumatic fever

There are additionally extraordinary coronary illness factors, from that most normal are recorded in the table beneath with their indication

TABLE - 2: FACTORS OF HEART DISEASE

Risk factors	Description	General Symptom
Age	Old people are more suffers from heart disease	Chest pain Shortness of breath Irregular heartbeat Fatigue Fainting Swollen feet
Sex	Males are at greater risk than females	
Family history	If relatives have heart disease the probability of a person to have cardiovascular disease is high	
Smoking	Heart disease higher in smokers than nonsmokers people	
Poor diet	Diet food is essential for development of heart	
Blood pressure	Blood pressure can effect in narrowing hardening arteries, as well as thickening blood vessels[1], [2]	
High blood cholesterol levels	It increases formation of plaques	
Diabetes	It is the disease as a result of sugar in our body	
Obesity	Overweight body is one of the cause for heart diseases	
Physical inactivity	Physical activity helps heart to function properly	
Stress	Damage arteries	
Poor hygiene	It increases heart disease	

III. RELATED WORK

Various works have been done identified with sickness forecast frameworks utilizing distinctive information mining strategies. Also, the AI calculations in clinical focuses by numerous creators, aiming to accomplish a precisely programmed finding of infections and better dynamic in clinical principles.

Marjia et al.[3] Propose coronary illness expectation utilizing KStar, J48, SMO, and Bayes Net and Multilayer perceptron using WEKA programming. Because of execution from various factors, SMO (89% of exactness) and Bayes Net (87% of accuracy) accomplish ideal performance than KStar, Multilayer perceptron, and J48 procedures utilizing k-crease cross-validation. The actual performance achieved by those calculations is as yet not palatable. So that if the presentation of exactness is improved more to give hitter choice to analysis infection.

Azam et al.[4] The paper depicts the programmed conclusion of coronary supply route sickness (CAD) patients utilizing streamlined SVM. These SVM boundaries are advanced to improve the precision of forecast, which gives an exactness of 99.2% utilizing k-overlay cross-approval. The paper serves to determine infection at the beginning phase and to lessen the expense. The equivalence acquired is a great idea to foresee if the individual has a coronary illness or not.

The Shailendra Narayan Singh et al. [6]Prediction of heart illness utilizing information mining methods, the paper portrays distinctive classifier benefits and faults for information arrangement and information extraction to execute calculations are generally helpful in wellbeing associations.

The AbhishekTaneja.[7] Heart illness expectation framework utilizing information mining strategies and diverse directed Machine learning calculations like J48, Naïve Bayes, and Multilayer insight inside WEKA Machine learning programming just as 10-overlap cross approval is applied. J48 outflanked Naïve Bayes classifier and Neural Networks when the prevalence relies upon the informational collection nature.

Sanavar et al.[8] The depiction of a review paper on coronary illness forecast. It portrays the distinctive strategy furthermore, how proposed methods are executed. It likewise gives some review of coronary illness, just as the function of information mining in medical care habitats and how to apply or use information mining in a medical services association is clarified.

Dr. S. Seema et al.[9] The paper centers around strategies that can anticipate constant illness by mining the information containing chronicled wellbeing records utilizing Naïve Bayes, Decision tree, Backing Vector Machine (SVM). A near report is performed on classifiers to quantify the better presentation at an impressive rate. From explore, SVM gives the most noteworthy exactness pace of 95.556% if there should arise an occurrence of coronary illness while for diabetes, Naïve Bayes gives the most remarkable precision of 73.588%.

The Rajendra Acharya et al. [10] Describe PC helped to find of the diabetic subject by pulse inconstancy signals utilizing discrete wavelet change technique using extraordinary classifiers that incorporate Decision Tree (DT), K-Nearest Neighbor (KNN), Naïve Bayes (NB), and Support Vector Machine (SVM). The normal exactness acquired is 92.02% by utilizing the DT inside ten times cross-approval. The processed accuracy is significant for forecast, notwithstanding, it isn't sufficient varying.

The Curtis Langlotz et al.[11] The paper centers around Bone Tumor Diagnosis Using a Naïve Bayesian Model Of Segment and Radiographic Features. The paper is based on two Naïve Bayes precision techniques, one is Naïve Bayes essential model, and the second is Naïve Bayes differential exactness. Utilizing Naïve Bayes Primary exactness, 62% of accuracy is accomplished, and using Naïve Bayes Differential exactness, 80% of precision is completed. The impediment of the paper is that the identity isn't sufficient for a superior choice.

The Vidya K. Sudarshan et al.[12] The paper is centered around applying higher-request spectra for the portrayal of

coronary conduit sickness utilizing electrocardiogram signals. The article uses a K-Nearest Neighbor and Decision Tree. The execution of exactness determined by these calculations is 98.17%, and 98.99% are gotten separately. The measures utilized in the paper give a better consequence of the presentation of exactness to portray the coronary supply route illness.

The Ashok Kumar Dwivedi.[13] The paper point is to assess various AI procedures for coronary illness expectation utilizing ten times cross-validation. The paper uses multiple calculations like Naïve Bayes, Classification Tree, KNN, Logistic Regression, SVM. Furthermore, ANN and the gives 83%, 77%, 80%, 85%, 82%, 82% and 84% separately. The Logistic Regression gives better exactness contrasted with different calculations.

The EmranaKabirHashi,[14] A specialist clinical choice emotionally supportive network to anticipate infection utilizing order methods. They give 90.43% and 76.96% exactness separately. C4.5 Decision Tree gives better precision contrasted with KNN and helps for clinical choice help framework.

HuseyinPolat et al.[15] The paper is utilized to Diagnosis of Chronic Kidney Disease dependent on SVM by highlight choice strategies. The article gives 98.5% most high precision accomplished by FilterSubsetEval with Best First. This paper's accuracy is critical to the conclusion of constant kidney infection for a better dynamic in clinical focuses.

MeghaShahi et al.[16] The paper's goal is a Coronary illness expectation framework utilizing information mining methods and WEKA programming for the programmed conclusion of illness and giving characteristics of administrations in medical services communities. The paper used different calculations like SVM, Naïve Bayes, Association rule, KNN, ANN, and Decision Tree. In some writings, the report clarifies that SVM's compelling and proficient precision is about 85% compared to other information mining calculations.

The Priti Chandra et al.[17] The paper depicts the Computational Intelligence Technique for Early Diagnosis of Coronary illness utilizing WEKA and 10-Fold cross-approval. The calculation used in this examination paper is Naïve Bayes, which gives 86.29% exactness. The exactness acquired is acceptable, yet there isn't agreeable to the programmed finding of heart sickness.

The Tapas RanjanBaitharu.[18] The paper's focal point is on the Analysis of Data Mining Techniques for Healthcare Choice Support System Using Liver Disorder Dataset. The paper utilized different calculations; those are J48, ZeroR, Multilayer Perceptron, IBK, Naïve Bayes, and VFI. The exactness acquired by calculations are 68.9%, 57.971%, 71.59%, 62.8986%, 55.3623% and 60.2899% separately. The estimation of exactness is better; however, not agreeable for the medical services dynamic framework.

The Syed Muhammad Saqlain Shah et al.[19] The paper is proposed for the investigation of Heart Disease Diagnosis dependent on highlight extraction utilizing K-Fold cross-approval. The calculation used by this paper is SVM, which gives 91.30%. The measure of exactness is better for the forecast of coronary illness and programmed analysis of the sickness.

Muhammad Saqlain et al.[20] The paper is centered around recognizable proof of Heart Failure by utilizing unstructured Data of Cardiovascular Patients. The paper utilizes Logistic Regression, Neural Organization, SVM, Random Forest Decision Tree, and Naïve Bayes. The calculations accomplish exactness of 80%, 84.8%, 83.8%, 86.6%, 86.6% and 87.7% separately for each individuals calculation. Gullible Bayes gives the highest exactness contrasted with other calculations.

You are comparing the accuracy performance of algorithms from related work.

TABLE - 3: COMPARISON OF ALGORITHMS USED IN VARIOUS RELATED WORK

Year	Author	Purpose	Techniques used	Accuracy
2016	Ninja et al. [3]	Heart disease prediction using WEKA tool and 10-Fold cross-validation	KStar	75%
			J48	88%
			SMO	89%
			Bayes Net	87%
			Multilayer Perceptron	90%
2017	Azmi et al. [4]	Automatic diagnosis of heart disease using K-Fold cross-validation method	Optimized SVM	99.2%
2018	Cauli et al. [5]	Propose application of knowledge discovering process on prediction of stroke patients	ANN	81.82% for training dataset
				85.9% for test data set
			SVM	80.38% for train data set
				84.26% for test data set
2011	AlHaideriTaneja [7]	Heart disease prediction system using data mining techniques and different supervised Machine learning algorithms	J48	95.56 %
				92.42 %
			Multilayer perceptron	91.89 %
	Rajendra	Computer-aided diagnosis of the	Decision Tree	

IV. HEART DISEASE DATA MINING

Information mining is significant in medical services associations for mechanizing the framework to improve the associations' working conditions. Information mining assists with giving improved characteristics of administrations and decrease costs. Today in medical services habitats, a vast measure of information is put away electrically, making it hard to examine customarily. Utilizing advancements to explore the massive report in data sets or other data vaults is essential to spare people's lives. A few calculations used for expectation in medical services are portrayed as follows.

A. Decision Trees

A choice tree is of managed learning calculation classifier, which is easy to comprehend and decipher. It bargains with both mathematical and absolute informational indexes. The choice tree would appear that tree structure, which comprises inner hubs, branches, and leaf hubs. Each chapter speaks to property estimations of a given dataset. Internal corners Donets a test on a given ascribes, and leaf hubs show the anticipated class or demonstrate the result results. The arrangement rule begins from the root hub to leaf hubs

to view the visionary quality and standards. The most well-known utilized choice tree calculations are a CART, ID3, C4.5, J48, and CHAID are significant in the expectation of diseases[21][3].

B. Naive Bayes Classifier

Credulous Bayes is utilized for characterization, which depends on Bayes' hypothesis. The events of specific highlights of a class are free of the presence or nonattendance of other highlights as indicated by Naïve Bayesian classifier hypothesis. It is a robust classifier for an expectation of coronary illness. Innocent Bayes is utilized for figuring the back likelihood of each class, which is contingent upon group informational collections. The condition is given as follows.

$$P(C|X) = \frac{P(X|C) * P(C)}{P(X)}$$

Where X is the example to be anticipated, and C is the class esteem, for instance. The above-given recipe or condition assists with deciding the class in which highlight expected to categorize[6]

C. Support Vector Machine

SVM is a strategy for grouping that manages both straight and nonlinear informational collections. SVM processing arrangement by augmenting the edge of the hyperplane that isolates the two classes. SVM is utilized portions capacities like the quadratic, polynomial, spiral premise, and so on to characterize the case.

V. DATA SET

Starlog coronary illness dataset taken from UCI machine learning lab is used for this examination paper. This information base contains 13 ascribes.

Property Information included are[22][1]:

- ✓ age
- ✓ sex
- ✓ chest torment type (4 qualities)
- ✓ resting circulatory strain
- ✓ serum cholesterol in mg/dl
- ✓ fasting glucose > 120 mg/dl
- ✓ resting electrocardiographic outcomes (values 0,1,2)
- ✓ the greatest pulse accomplished
- ✓ practice instigated angina
- ✓ old peak = ST melancholy actuated by practice relative
- ✓ to rest
- ✓ incline = the slant of the pinnacle practice ST section
- ✓ number of significant vessels (0-3) hued by fluoroscopy
- ✓ thal: (3 = typical, 6 = fixed deformity, 7 = reversible deformity)

VI. PROPOSED METHODOLOGY

The proposed philosophy's principal target is to foresee coronary illness for an early programmed finding of the sickness inside recovery to bring about a brief timeframe [23]. The Proposed methodology assumes essential parts for medical care specialists to treat their patients dependent on precise dynamics and give administrations characteristics to the individuals. Moreover, the proposed philosophy is basic in medical services Organizations with specialists with no more information and aptitude. One of the fundamental impediments of an existing procedure is the capacity to give precise outcomes as required. This framework use information related to mining procedures and machine learning calculations J48, Naïve Bayes, and Support Vector Machine, with k-crease cross-approval to foresee the event of coronary illness. It utilizes diverse clinical traits that are more important, such as age, sex, circulatory strain, cholesterol, blood sugar, and pulse; a portion of the traits are incorporated to distinguish if the individual has a coronary illness. Investigations of information sets are processing (executing) utilizing WEKA programming.

WEKA is open source programming that incorporates an assortment of AI calculations for the information mining undertakings. WEKA execution depends on Java code. WEKA contains devices for information preprocessing, relapse, grouping, characterization, affiliation decision, and representation significant in information mining errands.

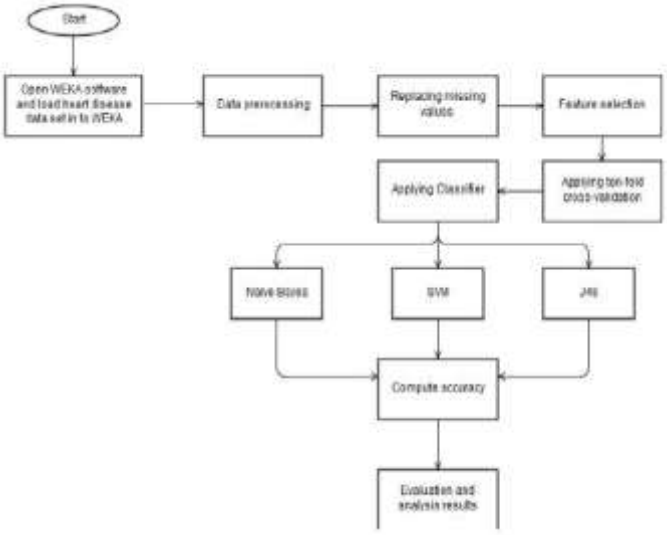

Fig - 1:flow diagram of the proposed methodology

VII. CONCLUSION

Coronary illness is the most widely recognized sickness that prompts passing in our reality, as indicated by the World Health Organization (WHO) reports, particularly in non-industrial nations. All clinical specialists don't have equivalent information and ability to choose, wherein a few specialists give a poor sensible choice that leads individuals to dangerous circumstances. To beat these issues, a forecast

of the event of illnesses is essential. One of the advantages of study papers is to improve the current approach for better dynamic by utilizing various calculations and highlight determination strategies. The proposed method uses J48, Naïve Bayes, and Support Vector Machine calculations for foreseeing the event of coronary illness for early programmed finding and quick timeframe recovery result that assists with giving the characteristics of administrations and decrease expenses to spare the lives of people.

VIII. FUTURE WORK

Decide the expectation execution of every calculation; what's more, apply the proposed framework for the zone it required. Utilize more effective element determination techniques to improve the precise performance of measures. Make the partners utilize this proposed strategy for making an alluring working condition that helps for better dynamic.

REFERENCES

[1] R. Rao, "Survey on Prediction of Heart Morbidity Using Data Mining Techniques," Knowl. Manag., vol. 1, no. 3, pp. 14-34, 2011.
[2] M. C. Staff, "Heart disease," Mayo Clinic. [Online]. Available: http://www.mayoclinic.org/diseasesconditions/heart-disease/symptoms-causes/dxc-20341558.
[3] 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.
[4] A. Davari Dolatabadi, S. E. Z. Khadem, and B. M. Asl, "Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM," Comput. Methods Programs Biomed., vol. 138, pp. 117-126, 2017.
[5] C. Colak, E. Karaman, and M. G. Turtay, "Application of knowledge discovery process on the prediction of stroke," Comput. Methods Programs Biomed., vol. 119, no. 3, pp. 181-185, 2015.
[6] M. Gandhi, "Predictions in Heart Disease Using Techniques of Data Mining," Int. Conf. Futur. trend Comput. Anal. Knowl. Manag., 2015.
[7] Gadde, S. S., & Kalli, V. D. R. Artificial Intelligence To Detect Heart Rate Variability. 10.33144/23939516/IJETA-V7I3P2
[8] S. Kiruthika Devi, S. Krishnapriya, and D. Kalita, "Prediction of heart disease using data mining techniques," Indian J. Sci. Technol., vol. 9, no. 39, pp. 21-24, 2016.
[9] Gadde, Sai & Kalli, Venkata. (2020). IJARCCE A Qualitative Comparison of Techniques for Student Modelling in Intelligent Tutoring Systems. 9. 75-82. 10.17148/IJARCCE.2020.91113.
[10] U. Rajendra Acharya, K. S. Vidya, D. N. Ghista, W. J. E. Lim, F. Molinari, and M. Sankaranarayanan, "Computer-aided diagnosis of diabetic subjects by heart rate variability signals using discrete wavelet transform method," Knowledge-Based Syst., vol. 81, pp. 56-64, 2015.
[11] B. H. Do, C. Langlotz, and C. F. Beaulieu, "Bone Tumor Diagnosis Using a Naive Bayesian Model of Demographic and Radiographic Features.," J. Digit. Imaging, 2017.
[12] U. R. Acharya et al., "Application of higher-order spectra for the characterization of Coronary artery disease using electrocardiogram signals," Biomed. Signal Process. Control, vol. 31, pp. 31-43, 2017.
[13] Ashok Kumar Dwivedi, "Analysis of computational intelligence techniques for diabetes mellitus prediction," Neural Comput. Appl., vol. 13, no. 3, pp. 1-9, 2017.
[14] E. K. Hashi, M. S. U. Zaman, and M. R. Hasan, "An expert clinical decision support system to predict disease using classification techniques," 2017 Int. Conf. Electr. Comput. Commun. Eng., pp. 396-400, 2017.
[15] H. Polat, H. Danaei Mehr, and A. Cetin, "Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods," J. Med. Syst., vol. 41, no. 4, 2017.
[16] M. Shahi and R. Kaur Gurm, "Heart disease prediction system using data mining techniques," Orient. J. Comput. Sci. Technol., vol. 6, no. 4, pp. 457-466, 2013.
[17] N. O. Fowler, "Diagnosis of Heart Disease," vol. V, no. March, pp. 1-7, 2012.
[18] Gadde, S. S., & Kalli, V. D. R. Descriptive Analysis of Machine Learning and Its Application in Healthcare. 10.33144/23478578/IJCST-V8I2P28
[19] Gadde, S. S., & Kalli, V. D. R. Medical Device Qualification Use. 10.17148/IJARCCE.2020.9410
[20] M. Saqlain, W. Hussain, N. A. Saqib, and M. A. Khan, "Identification of Heart Failure by Using Unstructured Data of Cardiac Patients,"2016 45th Int. Conf. Parallel Process. Work., pp. 426-431, 2016.
[21] T. Karthikeyan, B. Raghavan, and V. A. Kanimozhi, "A Study on Data mining Classification Algorithms in Heart Disease Prediction," Int. J. Adv. Res. Comput. Eng. Technol., vol. 5, no. 4, pp. 1076-1081, 2016.
[22] A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease, "Neural Comput. Appl., pp. 1-9, 2016.
[23] K. Buchan, M. Filannino, and Ö. Uzuner, "Automatic prediction of coronary artery disease from clinical narratives," J. Biomed. Inform., vol. 72, pp. 23-32, 2017.