International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-11 Issue-4, April 2022

Study on Predicting Heart Disease Diagnosis with Hybrid Machine Learning Techniques

Venkateswara Rao Cheekati, D. Natarajasivan, S. Indraneel

Abstract*: Machine learning can successfully forecast cardiac disease. The main benefit of these systems is their adaptability in non-linear contexts, allowing them to handle new data sets. Heart illness is the most common. We examined many indicators to better predict heart illnesses and also applied algorithms to forecast them. Modernity encourages us to be more active and fit, but it also pushes us to push ourselves harder and risk injury. These ecosystem-wide advancements have given bacteria, viruses, and other diseases a substantial new capability in this setting. Heart failure seems to be on the rise. Blood pressure, sugar, heart rate, and other markers are cardiovascular risk factors that cause blood arteries to be restricted or locked. Aneurysm, heart, or stroke. It can cause heart disease, vascular disease, CVA, cardiac death, and sudden death. Medical exams are used to diagnose various cardiac conditions, but the patient's family history and other factors should be considered. It's more tough to conclude for folks who don't get checked and have heart failure. Heart disease is one of the most common ailments nowadays, and early detection is critical to saving lives. The goal of this article is to improve accuracy, reduce training time, and reduce unknown cases by evaluating multiple classifiers on the data set to discover optimal HD attribute configurations. The K-Nearest Neighbor (K-NN), Naive Bayes, and SVM were compared to represent, JR and Adrost Decision Tree (JRandom), in order to assess the potential

***Keywords*: About Four Key Words Or Phrases In Alphabetical Order, Separated By Commas.**

I. INTRODUCTION

In the present day, cardiac insufficiency is a product of high blood pressure, hormone fluctuation, as well as excessive cholesterol and sugar levels, and a change in lifestyle is a contributing cause. The disease has been identified for a long time, and a variety of analytics of data have been developed to aid healthcare professionals recognize certain signs. Additional care may assist prospective patients deal with HD treatment.

The study revealed a range of extra inspections to minimize the disease's behavior and effective ways for early identification, including a risk factor for advancement in research. The SVignerba study reported on the application of K-NN and other other machine learning (ML) algorithms such as the Stochastic Gradient Boost, Naives Bayes, and Support Vector Machines (SVM), on the HD data set [They produced good results on JRK classification and prediction (SVM) [1].

Due to the fact that HD is too hard to diagnose and the right design tests are usually performed; as well as the suggested tests, logistic naive event classification, which achieved a detection rate of 82 percent and precisely predicted when patients would be admitted and three other machine learning algorithms testing showed the best result [2]. Logistic naive event categorization was more accurate due to data acquisition and results could be achieved than expected, in particular in view of the relatively high false positive and low negative rates [3].

In clinical investigators began [4] by using neural networks (particularly to assess cardiac disease) as part of the Backpropagation lp Multilator to forecast the existence of the disease [5] as clinical researchers showed [6] that multilayered perceptional integration (multi-invader network) methods are effective in the prognosis of disease) [7]. Experimental data indicated that Random Forestifier might increase the assessment of HD risk. Several statistical data mining and neural network classification approaches were combined in a mixed method of prevision methodology (7). The data qualities such as neighboring K-Nearest trees, genetic algorithms, and Naïve Bayes give good classifications to detect a probable heart illness.

However, the classification of the risk by classifications such as Bayesian means that sophisticated Bayesian K-Ne classifications like Genetic Recalibler and Naive Bayes are right, but cannot identify or identify an individual. The Bayesian modal information CADIS, showing that [8] was presented and found to cut by half the time necessary for HD predictions in the data set, while the methodology known as HD ADVIS exhibited a fifty percent shorter predictive time required for HD datasets [9].

Authors in [10] used optimization of particulate swarms to enhance efficiency and also highlight their suitability to the efficiency of machine learning and machine learning systems in a comparative analysis of the diagnosis of cardiopathy as a survey paper [11]. - actual- a patient arteria (Coronary) Classification (Algor) Perfusion CoaClass score has been rendered on a regular basis with the objective of fast screening, notification, cardiac (perfusion) diagnosis and medical review (clas and advance) patients, using a CoAl clinical effectiveness of more than 85 per cent [12]. The authors [13] employed the classification algorithm of Nave Bayes to diagnose HD patients and presented a Heart Disease Prediction System (HDPS) that is based on study of numerous algorithm parameters.

Manuscript received on 08 September 2021.
Revised Manuscript received on 07 April 2022.
Manuscript published on 30 April 2022.

* Correspondence Author

Venkateswara Rao Cheekati*, Research Scholar, Department of Computer Science and Engineering, Annamalai University, Chidambaram (Tamil Nadu), India. E-mail: chvraograce@gmail.com

S. Indraneel, Department of Computer Science and Engineering, Acharya Nagarjuna University, Guntur (Andhra Pradesh), India. E-mail: natarajasivan@gmail.com

Dr. D. Natarajasivan, Professor, Department of Computer Science and Engineering, Acharya Nagarjuna University, Guntur (Andhra Pradesh), India. E-mail: sreeram.indraneel@gmail.com

II. BACKGROUND WORK

The prediction of HD with the K-median clustered algorithm has been demonstrated [14] where authored authors proposed a successful prediction method for hybrid cardiac diseases using 14 out of the 74 attributes of UCI's Heart Disease Collection and the parameters of age [15], weight, sex, blood pressure and cholesterol as predictions in our paper [16]. for the investigation of heart rate variability introduced a new paradigm for non-linearity, combined with many classification techniques [17]. In the authors suggested a fair approach to predict the probability that HD would be acquired under categorisation determination law [18]. has suggested a strategy to categorizing the HD dataset with the use of 13 heart disease prediction medical features of neural networks (NN), with preliminary results indicating the higher performance of the proposed algorithm Other prediction algorithms [19]. highlighted the usage of Artificial Neural Network classification (ANN) systems for predicting an HD dataset by employing a back-stage method for train the network with 13 clinical inputs and predicting the 95 per cent accuracy of cardiac disease absence or existence [20]. In numerous precursor approaches were introduced using UCI Laboratory data, using an algorithm of 86.8 per cent F-size, for the search of templates such as Decision trees, Neural networks, Rough sets, SVM and Naive Bays [21]. In the Artificial Network proposed a method of categorizing of carotid artery stenting (CAS) illness (ANN). 317 Taiwanese patients with NHIRD were utilized for training and checking for an ANN model with a cloning rate of 13 risk
factors and a frequency of serious cardiovascular adverse effects ANN model has an input of 13 risk variables (MACE) [22]. Their model achieved 89.4 percent sensitivity and 82.5 percent accuracy Recommended HD data set classification utilizing the voting methods for classification and estimate [23].

In the authors devised an approach that uses input to medical test results, extracted from the Probabilistic Principal Component Analysis (PPCA) sub-sample of reduced dimension attributes, and used the UCI data set to identify the heart illness. On average, the recommended methodology produced 86.43% of the data set [24]. suggested a Coronary Cardiac Disease Classification Model, employing Vector Support (SVM) and Artificial Neural Network (ANN), and proposed a reasonable, precises and rapid implementation of the Coronary Heart Database and UCI Machine Learning Database Medical Policy.

The authors [25] presented a multilayer Perceptron prediction approach on neural network heart disease [26]. The NN system provides 13 features to evaluate the incidence or lack of cardiac disease in patients with 98 percent accuracy via back propagation [27]. More recently, the authors of have used a collection of CAD data including workplace and environmental characteristics and additional clinical characteristically characteristics, and the proposed function selection system has achieved a precision of 81.23 percent using the SMOTE and XG Boost classificators [28]. The authors of compared and noted their weakness in previous research by many scientists on the collection of information and the presentation of professional frameworks for the diagnosis of coronary artery disease [29]. In the authors extended the dataset Z-Alizadeh Sani, which now includes 303 records with 54 attributes [30], to present a new way of data discretion selection which tackles the complexity of the CAD prediction [31]. In the Discernible Wavelet Transform (DWT) has been coupled with the novel 1D-HLP (1-D-HLP) approach for the automatic diagnosis of arrhythmias with 95% accuracy when 17 arrhythmias are identified using an MIT-BIH arrhethmia CEG data set [32]. suggested a strategy to non-linear morphological categorization with a vote scheme on atypical cardiac morphologies [33].

32.00 milliseconds and above (eLONG) for the quantification of cardiac failure [34]. This new type of ECG test was able to detect highly sensitive and exact arrhythmias at 99.37% and 94.62% [35]. Sonam Nikhar and others have carried out an important work using a paper aiming to anticipate the application of machine learning algorithms for future cardiac diseases [36].

The objective of this work is to provide an extensive introduction to naïve Bays and to the complete classification of the following extended paraphrase [37]. Some research has shown that the execution and proof of high-level data technology Superior results in a successful predictive model forecast that previously used the Decision Tree [38]. Id Gava Kokula, Devkar and Ghaneelare are establishing a project that anticipates cardiomyopathy prognosis [PhD] MLP (also known as multilayer perceptrons) have been used in this trial in the formation and evaluation of multilayer perceptrons (MLP). The graph above displays the main elements of the data flow structure of this paradigm. It is divided into numerous levels, one for input, one for output and several layers between the two. All nodes in the input layer are linked to the end nodes through these.

III. COMPARATATIVE ANALYSIS

In this paper we will first give a comparative examination of the classification of the HD dataset using several classifiers used in biological companies.

These classifiers were cross-validated with the 10-times approach and the accessible data set was used to learn the output of the classifier. In the context of a sensitivity analysis on which the J48 classifier prune factor depends on identifying the classification precision of this classifier as well as possible better classification a contribution to this paper would be made.

Finally, we can use Classifier Subset Evaluator to expand the set of attributes from which we can make HD devices predictions for all classification algorithms to the only ones that can successfully identify, with the only ones that can be used to measure their accuracy and to achieve better results for systems that are precise in HD d Men and women have heart failure as well as contributing to their top cause of mortality.

The early signs and effects of heart failure are essential for quick actions, and it will save lives for people with ongoing cardiac arrest. The risk of recovery increases as soon as medical care is provided.

This study presents mainly several strategies for research on heart attacks that are examined and appraised. Current prediction performance can be increased, which therefore demands the creation of new, inventive algorithms and ways to manage future difficulties. The analysis is shown in the table below.

YEAR	AUTHOR	PURPOSE	TECHNIQUE USED	ACCURACY
2016	SONAM NIKHAR [40]	Prediction of heard disease using Machine Learning Algorithm	Nave Bayee Classification Decision Tree	Decision tree has better accuracy when compared to Bayes Classification
2018	V.V Ramalingam [42]	Prediction of Heart disease using Machine Learning	Nave Bayes Classification SVM KNN Decision Tree Random forest Ensembled model	SVM has more accuracy than other techniques
2018	Aditi Gavhane [41]	Machine Learning Prediction of Heart Disease	Multi Layer perceptron Algorithm	MLP gives better accuracy
2019	Abhay Kishore1 [43]	Deep Learning Prediction of Heart Attack	RNN	RNN has 98% accuracy than others

IV. OVERVIEW OF THE PROPOSED SYSTEM

Disease prediction is well done by the implementation of the protocol that describes methods of study for the classification of patients' cardiovascular disease prediction model. The model provides the basis for cardiac failure predicting with any machine Methodology of learning. To be able to use a Hearts attack dataset algorithms during the preparation of classification process Evaluation of the Efficiency of Random Forest Decision Trees Using Naive Bays (Folds and Percentage Splits) Steps to validate the Qualified Classification Models Hearts Attack Prediction Using Random Forest Decision Trees and Inexperienced Bays Efficient machine teaching tactics for the identification of cardiac disease focus on machine teaching techniques used in certain Internet of Things in recent breakthroughs (IoT) [39]. The suggested approach starts with pre-processing of data. The property of numerous groups is used to determine if heart failure occurs or is not present [40]. If the patient has heart failure, it is set to 1, otherwise the patient is set to 0, which means that the patient cannot suffer from heart illness [41]. The evidence was pre-processed through the translation of patient records into diagnostic values. Data for 297 medical reports showed that 137 records had a value of 1, indicating the appearance of cardiovascular diseases, while the remaining 160 records had a value of 0, showing the absence of a heart ailment [42]. Naive Bayes, GLM, DL and SVM are utilized to meet the criteria of features for second stage selection and reduction [43]. The third stage consists of classification modelling and the results are presented following the use of the top performing classification approaches in machine learning.

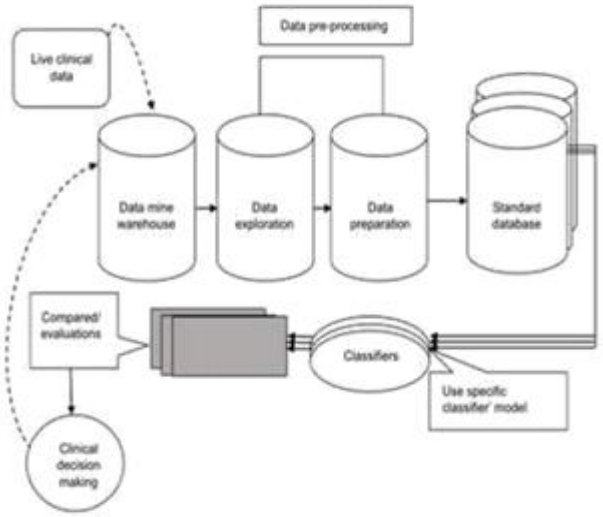

Fig 1:Methodology of Mining Heart Disease Data

The analysis approach for this study involves evaluations of the output of the three classification algorithms with cross-validation and percentage divisions. Cross-validation shall be performed by extracting research and testing information from the heart with multiple folds like ten and by remedying each fold in the data set for testing and training. The major aim of the research is to diagnose different cardiac illnesses and to take every conceivable precaution to prevent them at an affordable charge at an early stage.

1) To create a benchmark for predicting heart disease incidence.
2) Furthermore, this study attempts to find the ideal classification system for the risk of heart attack in a hospital.
3) This research is shown by the comparative review and examination at various assessment stages of three classification algorithms, namely Naive Bayes, Decision Tree and Random Forest.
4) While these are typical algorithms for machine learning, cardiac illness prediction is a critical task that demands the greatest possible accuracy.

The three algorithms are therefore analyzed at a number of stages and by means of a range of assessment methodologies. This will allow researchers and physicians to obtain a better knowledge of the situation and develop a method for the prediction of heart disease For the extension genetic algorithms and fuzzy logic approaches can be used to forecast heart illness, pick features of the genetic algorithm and categorize and predict the furious logic algorithm.

V. PROPOSED ARCHITECTURE

Regression classification and random wood classifications are commonly utilized to gain improved accuracy with other algorithms. The article shows five algorithms of classification used to generate the DM model. Cross-validation was employed ten times to sample data sets for preparation and research. The results and the accuracy of each experiment are evaluated using performance measures such as true positive rate, accuracy, F-measure, ROC field, Kappa statistics and a root-mean square error

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

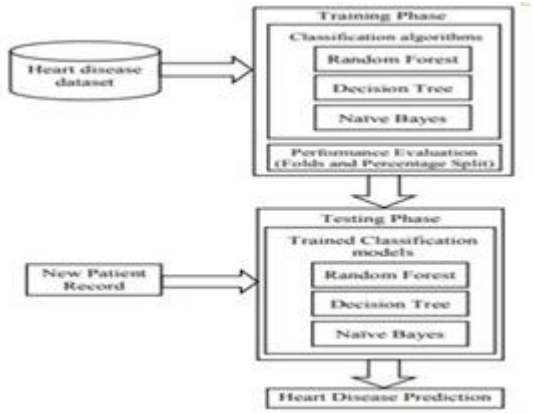

Fig 2: Proposed Methodology Architecture

The prediction of cardiac diseases can be done using a suggested technique that defines the research methods for developing a classification model required for the prediction of cardiac disease in patients. The model is a crucial procedure for predicting cardiovascular diseases utilizing all machine learning methods

VI. CONCLUSION

SVM, K-NN, and Naive Bayes are among the most extensively used algorithms in this work. To assess model accuracy, a comparative study is conducted. Pre-treatment via data affects accuracy by reducing a number of attributes. This is a CVD study. This paper proposed data mining for disease prediction. It attempts to define innovative data collection strategies from databases to help health practitioners. The output can be quantified by the time necessary to create the system's decision tree. The goal is to forecast the disease with the fewest attributes feasible. Many DSS can predict heart failure using various methods. Globally, heart disease is on the rise. Thus, it is critical to develop an automated system that accurately predicts cardiac diseases based on gender, age, and domain experience of experts in the area, while utilizing the most modern machine learning techniques available

REFERENCES

1. Khourdifi Y, Bahaj M. Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization. Int J IntellEng Syst. 2019;12(1):242–52.
2. Mohan S, Srivastava CTAG. Effective Heart Disease Prediction using Hybrid Machine Learning Techniques. IEEE Access. 2016.
3. Dai W, Brisimi TS, Adams WG, Mela T, Saligrama V, Paschalidis IC. Prediction of hospitalization due to heart diseases by supervised learning methods. Int J Med Inform. 2015;84(3):189–97.
4. Durairaj, M. Revathi, V., 2015. "Prediction Of Heart Disease Using Back Propagation MLP Algorithm". , 4(08), 235–239.
5. Gavhane A. "Prediction of Heart Disease Using Machine Learning" Second International Conference on Electronics, Communication and Aerospace Technology (ICECA); 2018. p. 1275–8.
6. Abdullah AS. "A Data mining Model for predicting the Coronary Heart Disease using Random Forest Classifier", Proceedings on International Conference in Recent trends in Computational Methods, Communication and Controls (Icon3c); 2012. p. 22–5.
7. BandarageShehaniSankethaRathnayakc ; Gamage UpekshaGanegoda , "Heart diseases prediction with Data Mining and Neural Network Techniques", 2018 3rd Int Conferee Convergence Technology (I2CT), pp.1–6.
8. Kelwade JP. "Radial basis function Neural Network for Prediction of Cardiac Arrhythmias based on Heart rate time series", Conference. IEEE First Int Conf Control MeasInstrum (CMI) (Cmi). 2016, 2016:454–8.
9. Anooj PK. Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules. J King Saud Univ Comput Inf Sci. 2012;24(1):27–40.
10. Krishnaiah V, Chandra NS. Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review. Int J Comput Appl. 2016;136(2):43– 51.
11. Fatima M, Pasha M. Survey of machine learning algorithms for disease diagnostic. J Intell Learn Syst Appl. 2017;9(01):1.
12. Otoom AF, Abdallah EE, Kilani Y, Kefaye A, Ashour M. Effective diagnosis and monitoring of heart disease. Int J Software Eng Appl. 2015;9(1):143–56.
13. Vembandasamy K, Sasipriya R, Deepa E. "heart diseases detection using naive Bayes algorithm", IJISET-international journal of innovative science. Eng Technol. 2015;2:441–4.
14. Malav A, Kadam K, Kamat P. Prediction of heart disease using k-means and artificial neural network as a hybrid approach to improve accuracy. Int J Eng Technol. 2017;9(4).
15. Lee HG, Noh KY, Ryu KH. "Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV,"Pacific-Asia Conference on Knowledge Discovery and Data Mining , Emerging Technologies in Knowledge Discovery and Data Mining; 2007. p. 218–28.
16. Tarle B. An artificial neural network based pattern classification algorithm for diagnosis of heart disease. Int Conf ComputCommun Control Automation (ICCUBEA). 2017:1–4.
17. Saxenab K, Purushottam RS. Efficient Heart Disease Prediction System. Procedia Comput Sci. 2016;85:962–9.
18. Al-milli N. Backpropogation Neural Network for Prediction of Heart Disease. J Theor Appl Inform Technol. 2013;56(1): 131–5.
19. Karaylan T, Kilic O. Prediction of heart disease using neural network. Int Conf Comput Sci Eng (UBMK) Antalya. 2017; 2017:719–23
20. Esfahani HA, Ghazanfari M. "Cardiovascular disease detection using a new ensemble classifier", IEEE 4th international conference on knowledge-based engineering and innovation (KBEI), Tehran, vol. 2017; 2017. p. 1011–4.
21. Cheng C, Chiu H. An artificial neural network model for the evaluation of carotid artery stenting prognosis using a National- Wide Database. 39th Ann Int Conf IEEE Eng Med Biol Soc (EMBC). 2017:2566–9. Uyar K, Ilhan A. Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks. Procedia Comput Sci. 2017;120:588–93.
22. Shah SMS, Batool S, Khan I. Muhammad Usman Ashraf, Syed Hussnain Abbas, Syed Adnan Hussain, "feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis". Physica A. 2017;482:796–807.
23. Radhimeenakshi S. Classification and prediction of heart disease risk using data mining techniques of Support Vector Machine and Artificial Neural Network. New Delhi: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom); 2016. p. 3107–11.
24. Sonawane JS, Patil DR. Prediction of heart disease using multilayer perceptron neural network. Chennai: International Conference on Information Communication and Embedded Systems (ICICES2014); 2014. p. 1–6. ICICES.2014.7033860
25. Nasarian E, Abdar M, Fahami MA, Alizadehsani R, Hussain S, Basiri ME, Zomorodi-Moghadam M, Zhou X, Pławiak P,
26. Acharya UR, Tan RS. Association between work-related features and coronary artery disease: a heterogeneous hybrid feature selection integrated with balancing approach. Pattern Recogn Lett. 2020;133:33–40.
27. Muhammad LJ, Garba EJ, Oye ND, Wajiga GM. "On the problems of knowledge acquisition and representation of expert system for diagnosis of coronary artery disease (CAD)", international journal of u-and e-service. Sci Technol. 2018;11(3): 49–58.
28. Alizadehsani R, Roshanzamir M, Abdar M, Beykikhoshk A, Khosravi A, Nahavandi S, Plawiak P, Tan RS, Acharya UR. hybrid genetic-discretized algorithm to handle data uncertainty in diagnosing stenosis of coronary arteries. Expert Syst. 2020.
29. Tuncer T, Dogan S, Pławiak P, Acharya UR. Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals. Knowl-Based Syst. 2019;186:104923
30. Kandala RN, Dhuli R, Pławiak P, Naik GR, Moeinzadeh H, Gargiulo GD, Gunnam S. Towards real-time heartbeat classification: evaluation of nonlinear morphological features and voting method. Sensors. 2019;19(23):5079.

31. Zomorodi-moghadam M, Abdar M, Davarzani Z, Zhou X, Plawiak P, Acharya UR. Hybrid real-binary particle swarm optimization for rule discovery in the diagnosis of coronary artery disease: Special Issue, Wiley online Library; 2019.
32. Abdar M, Książek W, Acharya UR, Tan RS, Makarenkov V, Pławiak P. A new machine learning technique for an accurate diagnosis of coronary artery disease. Comput Methods Prog Biomed. 2019;179:104992.
33. Pławiak P, Acharya UR. Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals. Neural Comput&Applic. 2019:1–25.
34. https://www.kaggle.com/johnsmith88/heart-disease-dataset.
35. https://archive.ics.uci.edu/ml/datasets/Heart+Disease.
36. Aggarwal CC. Data classification: algorithms and applications: CRC Press; 2014.
37. Cohen WW. Fast effective rule induction. Twelfth Int Conf Machine Lear. 1995:115–23.
38. Freund Y, Schapire R. A decision-theoretic generalization of on-line learning and an application to boosting; 1995.
39. Sethilkumar Mohan, ChandrasegarThirumalai, Gautam Shrivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," IEEE- journal IEEE-Access – 2019.
40. Sonam Nikhar, A.M. Karandikar" Prediction of Heart Disease Using Machine Learning Algorithms" International Journal of Advanced Engineering, Management and Science (IJAEMS) InfogainPublication,[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.
41. Aditi Gavhane, GouthamiKokkula, 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.
42. V.V. Ramalingam, AyantanDandapath, M Karthik Raja," Heart disease prediction using machine learning techniques: a survey", International Journal of Engineering & Technology, 7 (2.8) (2018) 684-687
43. 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.

AUTHORS PROFILE

Venkateswara Rao Cheekati, Research Scholar, Department of CSE Annamalai ,University, T.N , he is doing research in the area of Machine learning for Heart disease detection.

D.Natarajasivan, Assistant. Prof., Department .of CS,Annamialiah Univeristy He has be awarded PhD from Annammiala Univeristy, he has published various research paper and currently guiding Phd students for research.

Dr. Indraneel Sreeram, completed his Ph.d. in computer science and engineering from Acharya Nagarjuna University, Guntur in 2016. He is currently working as Professor in CSE department of St.Ann's College of engineering and technology, Chirala. His research interests are computer networks, Internet of things, Machine learning, Data Science