International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-7, Issue-6S,

A Systematic Literature Review on Cardiovascular Disorder Identification using Knowledge Mining and Machine Learning Methods

Syed Immamul Ansarullah, Pradeep Kumar

Abstract **Cardiovascular diseases are challenging to predict and diagnose due to the underlying dysfunctions associated with reflex mechanisms. Considering the mortality ratio and economy burden by the cardiovascular disorder, various researchers seek to diagnose this pernicious disease at its earliest by analyzing the healthcare data. In recent times, researchers made seminal contributions however, the unavailability of an extensive and fundamental article motivated us to prepare a literature review on a cardiovascular disease. We conduct a comprehensive database search between the years 2000 and 2017 using different keyword combinations to get distinguished articles about the disease. We provide descriptive insights to fill the uncovered research gaps. This paper attempts to uncover the state-of-the-art data mining approaches and tools that can be used to diagnose the cardiovascular disease at its initial. To our knowledge until now there is no competent and comprehensive article on cardiovascular disorder prognosis and identification using knowledge mining and machine learning approaches. The topic is diverse as well progressive hence demands additional research to understand newly identified discoveries about the disease.**

Index Terms**: *Data Mining, Association Rules, Cardiovascular Disease, Classification, Clustering, Healthcare Industry***

I. INTRODUCTION

Cardiovascular diseases (CVDs) are the disorders of heart and blood vessels that includes coronary artery disease, heart attack and heart murmurs, etc that knocks down a significant ratio of population. It is the growing socio-economic and public health problem that cause prolonged misery and disability. The Global Health Observatory (GHO) data show that CVDs are one amongst the fundamental causes of mortalities and disabilities in the world with 17.3 million deaths each year [52]. Below given figure1 shows the percentage of deaths by CVD across the globe during year 2014. The behavioral risk factors like cigarette usage, unhealthy diet, physical inactivity and the detrimental use of liquor are the primary cause of CVDs [54]. The effects of CVDs are divergent and its late identification can direct to severe complications including death. Despite being the foremost reason of mortality all over the world, CVDs are preventable so it is of central noteworthiness to perceive the condition with higher efficiency and accuracy in advance. Initial detection of CVD can help in recovering patients' health and reducing the mortality percentage. To diagnose CVDs in advance and diagnose the different researchers develop models on healthcare data using various knowledge mining and machine learning approaches as explained in below section III.

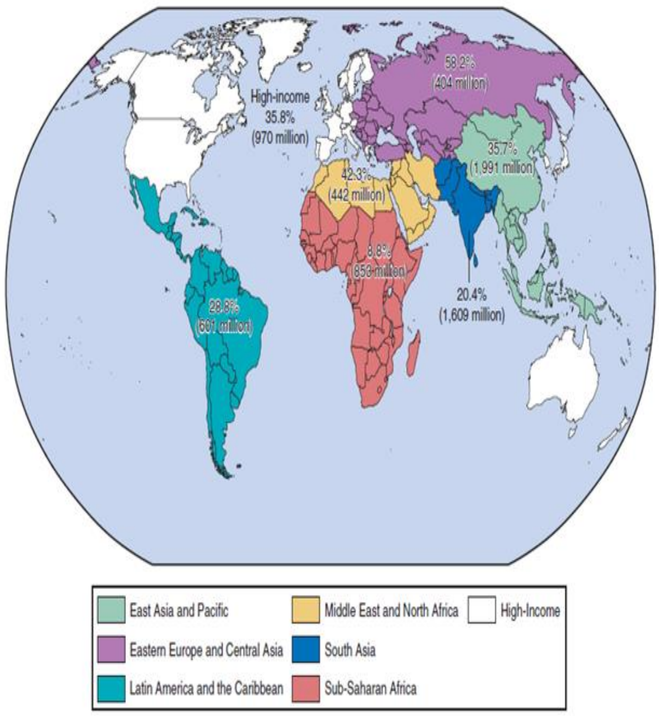

Figure1. CVD deaths as a percentage of all deaths in each region and total regional population during year 2014

II. RESEARCH METHODOLOGY

We reviewed the papers concerning machine learning and data mining approaches in diagnosing cardiovascular disease from the years 2000 to 2017. The period is significant because during this time the new data mining techniques and inventions that took place had incorporated in the healthcare industry. We used different search strings like "Cardiovascular disorder identification using knowledge mining and machine learning methods" etc in the following digital databases: IEEE Explorer, Google Scholar, Springer

Revised Manuscript Received on December 22, 2018.

Syed Immamul Ansarullah, Research Scholar Department of CS & IT Maulana Azad National Urdu University ,Hyderabad ,India

Dr. Pradeep Kumar, Associate Professor Department of CS & IT Maulana Azad National Urdu University ,Hyderabad ,India

Link, Elsevier, ACM Digital Library and MEDLINE. We found nine hundred thirty-two (932) papers that on further refining are filtrated to 69 articles on the topic basis for review purpose. Below given figure2 shows the pictorial representation of filtration of articles.

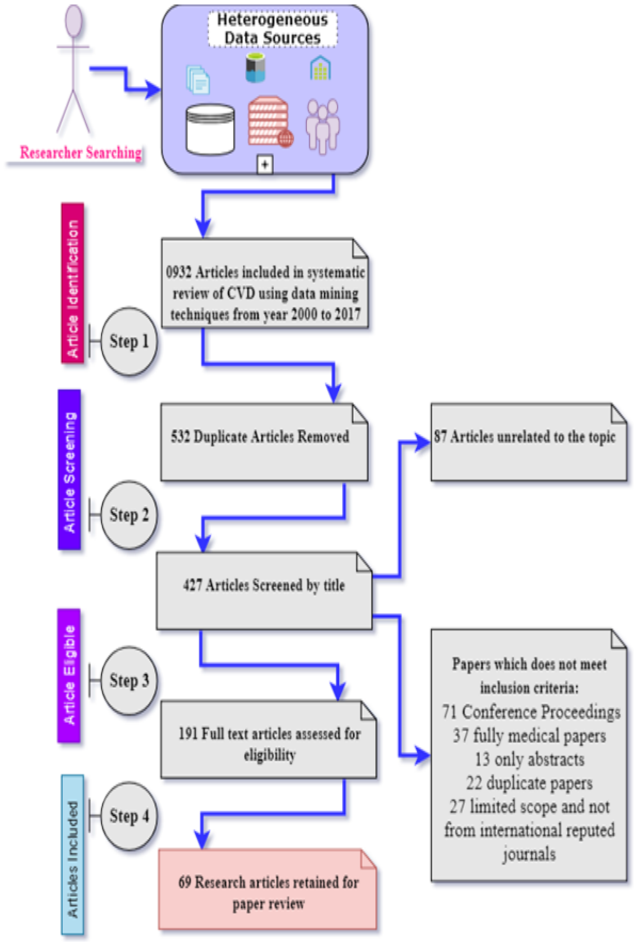

Figure2. Flow diagram of articles selected for review purpose

III. Challenges and Practices for Data Mining and Machine Learning in Healthcare Industry

The valuable inundated raw data in healthcare industries are languishing in warehouses. To get this data every researcher face different types of challenges. The challenges are, to collect applicable healthcare data from heterogeneous sources which are not free from losing data and a minor ratio of missing data can lead to severe complications and inaccurate decisions. The researchers have to keep under consideration the ethical, juridical, data ownership, data exposure, data preservation, information confidentiality and security issues. The researchers have to understand unstructured clinical observations in the proper context and extract potentially useful information from it. Below given subsections define the usage of different data science practices in healthcare industry especially in diagnosing cardiovascular diseases.

1. Classification Task

Different researchers use knowledge mining classification techniques on the healthcare data to extract valid and novel patterns for the early diagnosis of cardiovascular disease. The details of research done are summarized as follows:

Researchers like [1],[2],[3],[4] and [5] predicted and diagnosed cardiac disorder patients using K-Nearest Neighbor, Naïve Bayes, and Decision Tree algorithms. They build the respective heart disease models on different datasets with the varying number of attributes. The predictive models are validated using lift-chart and confusion matrix so as to determine which models provide right predictions for heart disease patients. Results show that K-NN outperforms other algorithms with the highest accuracy of prediction. Similarly, researchers like [6] build a heart disease classification model using Naïve Bayes algorithm. The researchers [7] predict and diagnose the cardiovascular disease by using variants of Naïve Bayes algorithm. The model performance is checked by accuracy, sensitivity and specificity measures. The results show that the heart disease predictive model outperformed than the traditional existing models. Researchers like [8],[9] and [10] used the decision tree and its variants to predict and diagnose cardiovascular disorder using minimum number of features. They build the respective heart disease models and evaluate the performance using measures like sensitivity, specificity, accuracy and error rate. The models performed well to identify the disease with optimal accuracies. Authors like [11], [12], [13], [14] and [15] diagnosed cardiac disorder using the clinical data of the patients. They used different feature selection techniques and variants of neural network algorithms to select the optimal and reduced set of attributes in cardiac disorder prediction. The data sets used in their respective research works are Cleveland database, Southwest, and Dajiang Hospital China and experimented on MATLAB and WEKA tool. The researchers [16] and [17] used different set of classification algorithms in combination with Genetic algorithm for the initial prediction of heart disorder. To get the most appropriate set of attributes for heart disease diagnosis they optimize the fitness function of the genetic algorithm. The outcome shows decision tree classifiers surpass other proposed classification algorithms after feature subset selection.

The researcher [18] identified heart disorder patients by building a predictive model. The model is built using the decision tree classification algorithm. Using Particle Swarm Optimization algorithm they select the optimal nine features for diagnosis and achieved an accuracy of 60.74%. Meanwhile the researchers like [19], [20], [21], [22] , [23] and [24] proposed the efficient ensemble classification methods on different datasets with varying number of instances for the CVD prediction and identification.

Researchers like [25] diagnose heart disease by using SMO, Naïve Bayes, C4.5 and AdaBoost algorithms. They used patients' laboratory and eco data to examine the set of features that play important role for coronary artery disease identification. The model build give an accuracy of 82% and to enhance it further researchers use the information gain method. Researchers like [26], [27], and [28] applied K-Nearest Neighbour algorithm and its variants for the early diagnosis of heart disease. The proposed models achieve the highest accuracy and reduced the complexity by using the feature selection techniques.

The authors [29], [30], [31], [32], [33], [34], [35], and [36] used numerous hybrid methods on their respective researches for the early identification of heart disease. Researchers imposed genetic search on various algorithms to select minimum crucial attributes for heart disease diagnosis. Evaluation outcome shows that the hybrid neural network algorithms and its variants outperform irrespective of the data set size. The researchers [37] and [38] developed the novel model for early diagnosis of heart disease by hybridizing fuzzy system with neural network and genetic algorithms. From the results, it is clear that the newly developed hybrid models diagnose the heart disease with great accuracy and outperform standard comprehensible methods. Researchers like [39], [40], [41], [42] and [43] build their respective heart disease models using the fuzzy methods on different datasets with the varying number of instances and attributes. Researchers obtained concise fuzzy rules that describe the disease with great accuracy.

2. Clustering

Clustering refers to the grouping of records into classes of similar objects. Many researchers evaluate the performance of different clustering algorithms like K-Means clustering, Mean-Shift Clustering, DBSCAN, and Expectation Maximization clustering using Gaussian Mixture Models, and Agglomerative Hierarchical Clustering for initial diagnosis of the cardiovascular diseases. Researchers like [44] implemented K-means clustering algorithm on UCI machine learning dataset to diagnose heart disease patients. The authors [45] diagnose cardiac disorder by means of Naïve Bayes classifier and K- Means clustering algorithm. The model achieves an accuracy of 84.5% by integrating K-Means and Naïve Bayes classifier.

3. Association Rules

Association rules have huge potential in the healthcare industry to diagnose diseases. Association approach is used to reduce the unrelated rules and get the rules with highest accuracy. Researchers like [46], [47], [48] and [49] predicted heart disease using Significant and interesting association rules with minimum support and confidence. Whereas the researchers [50] generated valid, accurate and optimal rules depending on the mentioned set of attributes. These optimal set of heart disease rules illustrate the patients risk level. To check the unbiased performance results they train and test the system using 10 fold cross validation methods which resulted in accurate and optimal results. Researchers [51] build a heart disease risk evaluation model using association rules. They choose the risk factors based on Apriori algorithms and test the performance of the model on WEKA tool. Similarly researchers like [52] extract the valid association rules for early diagnosis of heart disease. The search constraint is used to remove the banal rules and derive only valid rules that represent valuable medical knowledge.

We reviewed the papers concerning machine learning and data mining approaches in diagnosing cardiovascular disease from the years 2000 to 2017. After the literature review we come to know that different researchers had made seminal contributions using supervised and unsupervised tasks as shown in below given table1. Figure3 shows that the existing literatures mostly examines the Support Vector Machines, Naïve Bayes, Neural Network and Decision Trees algorithms with their respective variants for disease diagnosis and get an optimal accuracies of 93%, 94%, 95% and 95%. The figure3 also shows that researchers use the Association Rules and Logistic Regression for their study however they performed less. The papers published on cardiovascular disorder using knowledge mining and machine learning approaches across the years 2000 to 2017 is shown in Figure4. The distributions of papers show the growing interest of researchers and academicians in using data science in healthcare industry. We expect that analytics will continue be used in future on medical data to find the insights. We also checked the frequency of algorithms in the reviewed literature; the pictorial representation of it is shown in below given figure5. Among the articles included in the review we found that decision tree algorithms and their variants are mostly used while as bagging algorithms are less frequent.

After literature review we come to the conclusion that the final objective is to make up to date healthcare models capable of ensuring maximum patient satisfaction. We also found that respective research works improved the accuracy of CVD diagnosis as compared to the traditional systems. However from the contributions made by different researchers for cardiovascular disease identification an additional research is required to understand more completely the key factors that drag world to the disease.

IV. FINDINGS IN LITERATURE REVIEW

Numerous existing studies have examined cardiovascular disease prognosis and diagnosis however, a closer look to the literature reveals several shortcomings as described below:

1. Because of unseen and concise data set and over-fitting problems, all models lack in the generalization capability. However, to develop an efficient disease model and overcome these issues, the dataset should be complete and huge in size.
2. Some researchers' derived association rules from cardiovascular disease data set but, these rules are complex and large in nature which makes system slow and leads to inaccurate decisions.
3. Many researchers use decision trees in their respective works however the automatic splitting condition of decision tree algorithm on numerical medical variables makes experimental result interpretation more challenging and leads to wrong diagnosis for medical professionals. The medical society has standard splitting criterions which are universally acknowledged (high blood pressure, high cholesterol, male overweight, etc) hence decision tree algorithms should be trained on such cut-off values before applied on the medical dataset to get accurate predictions and restrain from wrong assumptions.
4. Several researchers included Naïve Bayes, Back Propagation and Genetic Methods for correct interpretation of CVD. However incorrect assumption nature of Naïve Bayes classifier degraded its performance; Slow convergence of the Back-propagation algorithm confines its usage for diagnosis of disease and the

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

difficulty in determining suitable fitness function of Genetic algorithm require lot of fitness assessments.

5. Many researchers used neural network algorithm in their respective studies to build the CVD models however, the error generated by the hidden neurons on output nodes degrade the logic potential of the neural network that results into the wrong prediction and decision making.

6. Researchers used cross validation and error rates for experiments and simulation purposes using different tools but each tool have complications accompanied with it. There are many mechanisms for prediction but all have limitations like Documentation for GUI is limited, scaling is a problem, can't handle Big Data, data preparation and visualization techniques offered might not be enough.

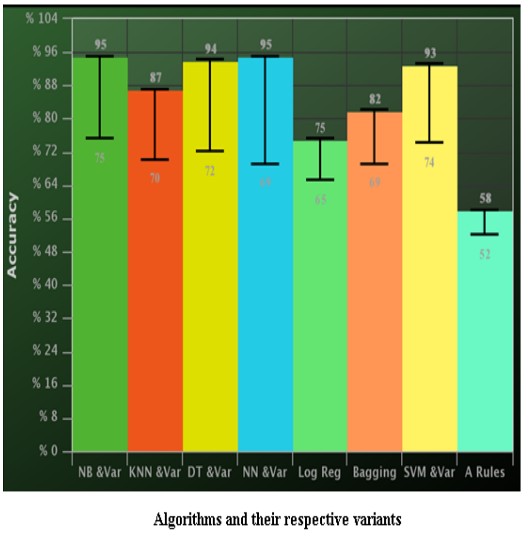

Figure3. Accuracies obtained by different algorithms and their variants for CVD diagnosis in reviewed literature during years 2000 to 2017

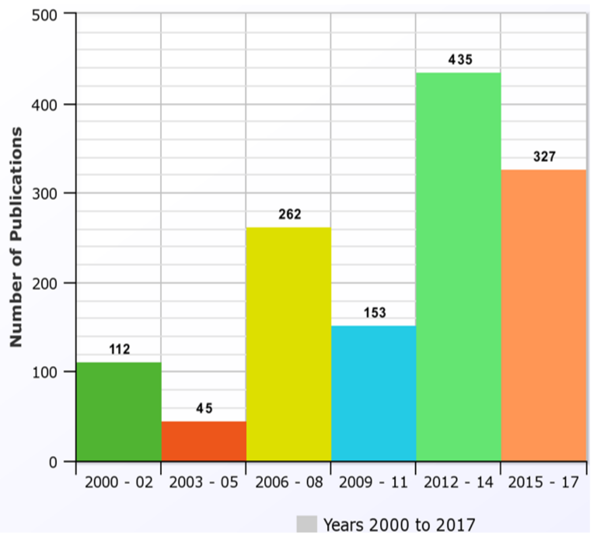

Figure4. Number of paper publications during 2000-2017aboutCVD using data mining techniques

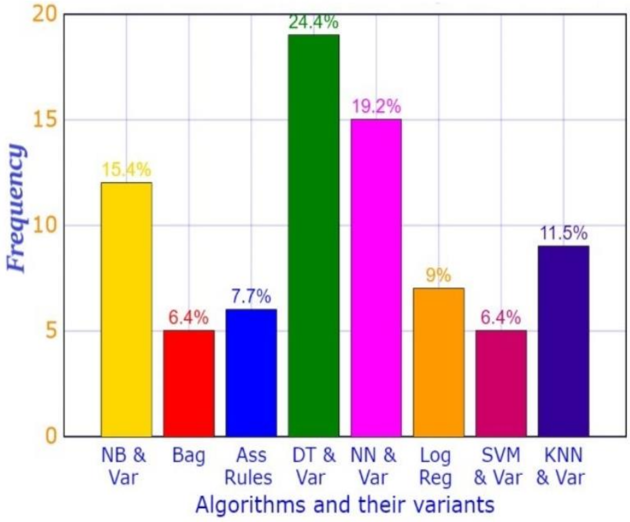

Figure5. Frequencies of different algorithms and their variants in reviewed literature for CVD diagnosis during 2000-2017

V. SUGGESTIONS

To improve healthcare systems and build effective models we need to follow the five core steps of machine learning workflow that are shown in below given figure6. Of course there are more situational steps required like project scoping, data wrangling, data preprocessing and Ensembling. However core steps are sufficient to build a successful machine learning model. Other suggestions that should be followed are mentioned as follows:

1. Instead of using conventional classification trees, researchers concerned in classifying CVD patients should use ensemble-based methods because it results in robust and optimal models.
2. We found that most of the predictive models can be extended by using real time datasets to get the accurate diagnosis in advance.
3. To get the disease relevant patterns and knowledge from sources like relational, transactional, object oriented, spatial and active databases most appropriate tactics and novel data demanding techniques should be used.

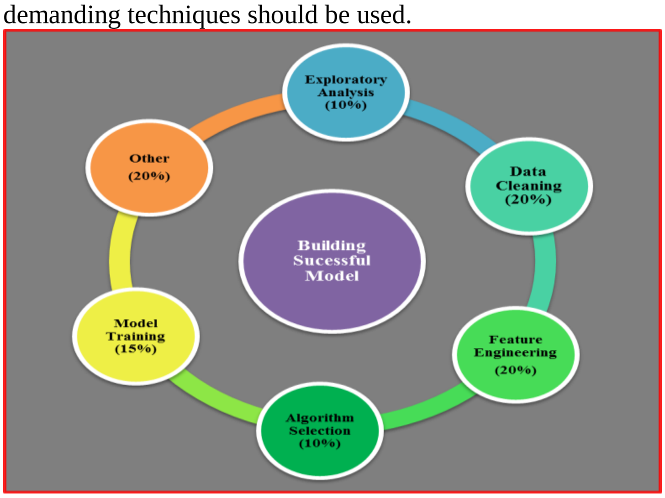

Figure6. Core Steps Required Building Successful Data Mining Model.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

VI. Conclusion

Knowledge mining and machine learning techniques are incorporated in the healthcare industry to find the insights from the inundated data. In recent times, researchers made decisive contributions on the cardiovascular disorder identification using knowledge mining and machine learning methods. We conduct a comprehensive literature review between the years 2000 and 2017 to find out the factors that drag world to cardiovascular diseases. We found that the behavioral risk factors like cigarette usage, unhealthy diet, physical inactivity and the detrimental use of liquor are the primary cause of cardiovascular diseases. Existing literature suggest that the effect of cardiovascular diseases are divergent and its late identification can direct to severe complications including death. We found that although cardiovascular disorders are the foremost reason of mortality all over the globe, yet it is preventable and controllable. To diagnose this disorder accurately in advance many researchers build models using the divergent data sets, different machine learning algorithms, various data mining approaches and numerous tools. We found that there is no single algorithm that produces best results for every dataset however, hybridization and ensemble methods show optimal results. The paper will benefit the medical academicians, practitioners, researchers who are interested in the areas of biomedical sciences. This paper includes to the research on cardiovascular disease and data mining, as to our knowledge until now there is no competent and comprehensive article on the disease however, the problem demands additional research to understand novel discoveries about the disease.

Table1 contributions made by different researchers in predicting and diagnosing CVD through data mining techniques

Researcher's / Year	Technique	Accuracy
Hall 2000	Naive Bayes	83.24
	K Nearest Neighbour	82.12
	Decision Tree	75.32
Colombet et al. 2000	CART	76.01
	Multilayer Perceptron	76.01
	Logistic Regression	78.01
Yan H et al. 2003	Multilayer Perceptron	63.06
Pavlopoulos, Stasis, and Loukis 2004	Decision Tree	91.01
Ture et al. 2005	Neural Network	95.01
Polat, Sahan, and Gunes 2007	Fuzzy-AIRS-k-Nearest Neighbour	87.01
Palaniappan S, Awang R 2008	Naive Bayes	95.01
	Decision Tree	94.93
	Neural Network	93.54
Tsipouras et al. 2008	Decision Tree with Fuzzy	80.01
Das et al. 2009	Neural Network Ensemble	89.01
Tu et al. 2009	Bagging Algorithm	81.41
Raj Kumar and Reena 2010	Naive Bayes	52.33
	K Nearest Neighbour	45.67
	Decision List	52.00
Dewan and Sharma 2015	Genetic algorithm and Back Propagation	87.23
Jabbar et al. 2015	Naïve Bayes	78.56
Olaniyi, Oyedotun, and Adnan 2015	Feed Forward Multilayer Perceptron	84.32
	Support Vector Machine	87.5
Ngueilbaye, Lei, and Wang 2016	Naive Bayes	91.42
	Support Vector Machine	89.56
C. Kalaiselvi 2016	K-Nearest Neighbor	96.21
Bialy et al. 2016	Naive Bayes	85.30
	Multilayer Perceptron	82.32
	Bayesian Net	79.98
	Decision Tree	81.11
	C4.5 Decision Tree	80.21
	Sequential Minimal Optimization	83.43
Arabasadi et al. 2017	Hybrid model of Neural Network and Genetic Algorithm	93.85

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Srinivas et al. 2010	Naive Bayes	84.14
	One Dependency Augmented Naive Bayes	80.46
Shouman et al. 2011	Information Gain	79.01
	Gini Index	82.33
	Gain Ratio	84.01
Chen et al. 2011	Learning Vector Quantization (LVQ) Neural Network	80%
Khemphila and Boonjing 2011	Multi-Layer Perceptron with Back-Propagation algorithm	80.17
Kumari and Godara 2011	Artificial Neural Network	80.06
	Support Vector Machine	84.12
	RIPPER	81.04
	Decision Tree	79.05
Soni et al. 2011	Classification based on Predictive Association Rules	52.32
	Classification based on Association Rule	58.28
	Weighted Association Classifier	57.75
	Classification based on Multiple Class Association Rules	53.64
Abdullah 2012	Particle Swarm Optimization- j48 Algorithm	60.74
Shouman et al. 2012	K-Nearest Neighbour	97.04
Purwanto et al. 2012	Bayesian Networks	60.02
	Multilayer Perceptron	64.01
	Radial Basis Function	54.40
	Logistic computational	60.30
Chaurasia 2013	CART	83.49
	ID3	82.21
	Decision Table	79.34
Alizadehsani et al. 2013	Bagging	70.35
	C4.5	74.09
Amin et al. 2013	Hybrid model Neural Network and Genetic algorithm	96.20
Chaurasia and Pal 2014	Decision Tree	83.78
	Naive Bayes	78.77
	Bagging	85.03
Srinivas K et al. 2014	Hybrid Model of Rough-fuzzy classifier	80.22
Elsayad and Fakhr 2015	Tree Augmented Naive Bayes	88.54
	Markov Blanket Estimation Naive Bayes	97.92

REFERENCES

1. I. Colombet, A. Ruelland, G. Chatellier, F. Gueyffier, P. Degoulet, and M. C. Jaulent, "Models to predict cardiovascular risk: comparison of CART, multilayer perceptron and logistic regression.," Proc. AMIA Symp., pp. 156–60, 2000.
2. V. Chaurasia and S. Pal, "Data Mining Approach to Detect Heart Diseases," Int. J. Adv. Comput. Sci. Inf. Technol. Vol, vol. 2, no. 4, pp. 56–66, 2014.
3. A. Ngueilbaye, L. Lei, and H. Wang, "Comparative Study of Data Mining Techniques on Heart Disease Prediction System: a case study for the," Int. J. Sci. Res., vol. 5, no. 5, pp. 1564–1571, 2016.
4. S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," 2008 IEEE/ACS Int. Conf. Comput. Syst. Appl., vol. 8, no. 8, pp. 343–350, 2008.
5. D. Masethe Hlaudi and A. Masethe Mosima, "Prediction of Heart Disease using Classification Algorithms," Proc. World Congr. Eng. Comput. Sci., vol. II, pp. 22–24, 2014.
6. M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Computational Intelligence Technique for early Diagnosis of Heart Disease," IEEE Int. Conf. Eng. Technol., vol. V, no. March, pp. 1–7, 2015.
7. A. Elsayad and M. Fakhr, "Diagnosis of cardiovascular diseases with bayesian classifiers," J. Comput. Sci., vol. 11, no. 2, pp. 274–282, 2015.
8. M. Shouman, T. Turner, and R. Stocker, "Using decision tree for diagnosing heart disease patients," Proc. Ninth Australas. Data Min. Conf., pp. 23–30, 2011.
9. V. Chaurasia, "Early Prediction of Heart Diseases Using Data Mining," Caribb. J. Sci. Technol., vol. 1, no. December, pp. 208–217, 2013.
10. M. S. Al-batah, "Testing the Probability of Heart Disease Using Classification and Regression Tree Model," Sci. Int., vol. 4, no. 11, pp. 1713–1725, 2014.
11. H. Yan, J. Zheng, Y. Jiang, C. Peng, and Q. Li, "Development of a Decision Support System for Heart Disease Diagnosis Using Multilayer Perceptron," IEEE Int. Symp. Circuits Syst., vol. 5, pp. 709–712, 2003.
12. N. Al-Milli, "Backpropogation Neural Network for Prediction of Heart Disease," J. Theor. Appl. Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.
13. K. U. Rani, "Analysis of Heart Diseases Dataset using Neural Network Approach," Int. J. Data Min. Knowl. Manag. Process, vol. 1, no. 5, pp. 1–8, 2011.
14. A. Chen, H. SY, P. Hong, C. Cheng, and L. EJ, "HDPS: Heart disease prediction system," 2011 Comput. Cardiol., pp. 557–560, 2011.
15. A. Khemphila and V. Boonjing, "Heart Disease Classification Using Neural Network and Feature Selection," 21st Int. Conf. Syst. Eng., no. 2007, pp. 406–409, 2011.
16. M. Anbarasi, E. Anupriya, and N. C. S. N. Iyengar, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm," Int. J. Eng. Sci. Technol., vol. 2, no. 10, pp. 5370–5376, 2010.
17. E. P. Ephzibah and V. Sundarapandian, "Framing Fuzzy Rules using Support Sets for Effective Heart Disease Diagnosis," Int. J. fuzzy Log. Syst., vol. 2, no. 1, pp. 11–16, 2012.
18. A. S. Abdullah, "A Data Mining Model to Predict and Analyze the Events Related to Coronary Heart Disease using Decision Trees with Particle Swarm Optimization for Feature Selection," Int. J. Comput. Appl., vol. 55, no. 8, pp. 49–55, 2012.
19. R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," Expert Syst. Appl., vol. 36, no. 4, pp. 7675–7680, 2009.
20. M. Kumari and S. Godara, "Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction," Int. J. Comput. Sci. Trends Technol., vol. 2, no. 2, pp. 304–308, 2011.
21. A. Aqueel and S. A. Hannan, "Data Mining Techniques to find out Heart Diseases : An Overview," (sem qualis) Int. J. Innov. Technol. Explor. Eng., vol. 1, no. 4, pp. 18–23, 2012.
22. M. . Jabbar, B. . Deekshatulu, and P. Chandra, "Alternating Decision trees for early diagnosis of dengue fever," in International Conference on Circuits, Communication, Control

and Computing, 2014, no. 5, pp. 322–328.
23. R. El Bialy, O. Karam, O. Karam, T. A. Randa, E. Bially, and M. A. Salama, "An ensemble model for Heart disease data sets : a generalized model," ACM, pp. 191–196, 2016.
24. M. C. Tu, D. Shin, and D. Shin, "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms," 2009 Eighth IEEE Int. Conf. Dependable, Auton. Secur. Comput., pp. 183–187, 2009.
25. R. Alizadehsani, J. Habibi, M. Javad, B. Bahadorian, and Z. Alizadeh, "A data mining approach for diagnosis of coronary," Comput. Methods Programs Biomed., pp. 1–10, 2013.
26. M. Shouman, T. Turner, and R. Stocker, "Using data mining techniques in heart disease diagnosis and treatment," Japan-Egypt Conf. Electron. Commun. Comput., pp. 173–177, 2012.
27. D. Chandna, "Diagnosis of Heart Disease Using Data Mining Algorithm," Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 5, pp. 1678–1680, 2014.
28. Kalaiselvi, "Diagnosis of Heart Diseases using Average K-Nearest Neighbor Algorithm of Data Mining," Int. Conf. Comput. Sustain. Glob. Dev., pp. 3099–3103, 2016.
29. A. Dey and J. Singh, "Analysis of Supervised Machine Learning Algorithms for Heart Disease Prediction with Reduced Number of Attributes using Principal Component Analysis," Int. J. Comput. Appl., vol. 140, no. 2, pp. 27–31, 2016.
30. A. Dewan and M. Sharma, "Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification," in International Conference on Computing for Sustainable Global Development (INDIACom), 2015, pp. 704–706.
31. E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, "Heart Diseases Diagnosis Using Neural Networks Arbitration," Int. J. Intell. Syst. Appl., vol. 7, no. 12, pp. 75–82, 2015.
32. B. V. Sumana and T. Santhanam, "Prediction of diseases by cascading clustering and classification," in 2014 International Conference on Advances in Electronics, Computers and Communications, ICAECC 2014, 2014, p. 8.
33. N. P. Waghulde and N. P. Patil, "Genetic Neural Approach for Heart Disease Prediction," Int. J. Adv. Comput. Res., vol. 4, no. 3, pp. 778–784, 2014.
34. S. U. Amin, K. Agarwal, and R. Beg, "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors," ICT 2013 - Proc. 2013 IEEE Conf. Inf. Commun. Technol., no. Ict, pp. 1227–1231, 2013.
35. Purwanto, C. Eswaran, R. Logeswaran, and A. R. A. Rahman, "Prediction models for early risk detection of cardiovascular event," J. Med. Syst., vol. 36, no. 2, pp. 521–531, 2012.
36. T. J. Peter and K. Somasundaram, "Study and Development of Novel Feature Selection Framework for Heart Disease Prediction," Int. J. Sci. Res. Publ., vol. 2, no. 10, pp. 1–7, 2012.
37. M. Abushariah, A. Alqudah, O. Adwan, and R. Yousef, "Automatic Heart Disease Diagnosis System Based on Artificial Neural Network and Adaptive Neuro-Fuzzy Inference Systems Approaches," J. Softw. Eng. Appl., no. November, pp. 1055–1064, 2014.
38. L. Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm," Int. J. Biol. Med. Sci. , vol. 3, no. 3, pp. 157–160, 2008.
39. A. Adeli and M. Neshat, "A Fuzzy Expert System for Heart Disease Diagnosis," Proc. Int. MultiConference Engineeers Comput. Sci., vol. I, pp. 1–6, 2010.
40. N. Bhatla and K. Jyoti, "A Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic," Int. J. Comput. Appl., vol. 54, no. 17, pp. 975–8887, 2012.
41. P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," J. King Saud Univ. - Comput. Inf. Sci., vol. 24, no. 1, pp. 27–40, 2012.
42. M. G. Tsipouras et al., "Automated diagnosis of coronary artery disease based on data mining and fuzzy modeling.," IEEE Trans. Inf. Technol. Biomed., vol. 12, no. 4, pp. 447–58, 2008.
43. K. Srinivas, G. R. Rao, and A. Govardhan, "Rough-Fuzzy Classifier: A System to Predict the Heart Disease by Blending Two Different Set Theories," Arab. J. Sci. Eng., vol. 39, no. 4, pp. 2857–2868, 2014.
44. S. B. Patil and Y. S. Kumaraswamy, "Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction," Int. J. Comput. Sci. Netw. Secur., vol. 9, no. 2, pp. 228–235, 2009.
45. M. Shouman, T. Turner, and R. Stocker, "Integrating Naive Bayes and KMeans Clustering with different Intial Centroid Selection methods in the diagnosis of heart disease patients," airccj.org, pp. 431–436, 2012.
46. M. Ilayaraja and T. Meyyappan, "Efficient Data Mining Method to Predict the Risk of Heart Diseases Through Frequent Itemsets," in 4th International Conference on Eco-friendly Computing and Communication Systems, ICECCS 2015, 2015, vol. 70, pp. 586–592.
47. J. Soni, U. Ansari, and D. Sharma, "Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers," Hear. Dis., vol. 3, no. 6, pp. 2385–2392, 2011.
48. C. Ordonez et al., "Mining Constrained Association Rules to Predict Heart Disease," in IEEE. Published in International Conference on Data Mining (ICDM), 2001, pp. 1–7.
49. K. Srinivas, G. R. Rao, and A. Govardhan, "Analysis of Attribute Association in Heart Disease Using Data Mining Techniques," vol. 2, no. August, pp. 1680–1683, 2012.
50. Purushottam, K. Saxena, and R. Sharma, "Efficient Heart Disease Prediction System," Procedia Comput. Sci., vol. 85, pp. 962–969, 2016.
51. M. A. Karaolis, J. A. Moutiris, D. Hadjipanayi, and C. S. Pattichis, "Assessment of the risk factors of coronary heart events based on data mining with decision trees," IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 3, pp. 559–566, 2010.
52. http://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
53. Mudasir M Kirmani and Syed Immamul Ansarullah. 2016, " Classification models on cardiovascular disease detection using Neural Networks, Naive Bayes and J48 Data Mining Techniques". International Journal of Advanced Research in Computer Science. Volume 7, No. 5, September-October 2016.
54. "WHO | Cardiovascular diseases (CVDs)," WHO, 2017.

AUTHORS PROFILE

Syed Immamul Ansarullah completed his BCA from Kashmir University and MCA from BGSBU. He also completed MPhil from APSU Rewa, M.P. He is currently pursuing PhD from the department of CS&IT, MANUU, Hyderabad. He has published several conference and journal papers in national and international repute.

Dr. Pradeep Kumar is an Associate Professor in the Department of Computer Science & Information technology at Maulana Azad National Urdu University, Hyderabad (Telangana State). He received his Master's degree in Computer Technology and Applications from Delhi Technological University. He completed his Ph.D. from the University School of Information & Communication Technology (USICT), Guru Gobind Singh Indraprastha University (GGSIPU), Delhi. His research interests include software reliability engineering, models for software metrics, machine learning, neural network modeling and soft computing. He has more than 25 publications in journals of international repute including national journals, conferences and proceedings of the international conferences. He is a Member of Association for Computing Machines (ACM), India, Member of Computer Science Teachers Association (CSTA), USA, Senior Member of International Association of Engineers (IAENG), Member of International Association of Computer Science and Information Technology (IACSIT), Singapore and Senior member of Universal Association of Computer and Electronics Engineers (UACEE). He is a member of editorial board for various national and international journals in the field of software engineering and program committee member/reviewer for several international conferences.