INTERNATIONAL JOURNAL OF RESEARCH IN MEDICAL SCIENCES & TECHNOLOGY

e-ISSN:2455-5134; p-ISSN: 2455-9059

EMPLOYABILITY OF SELECTED DATAMINING ALGORITHMS IN THE EARLY PREDICTION OF CORONARY DISEASES

Prachi Juneja

Sri Guru Gobind Singh College Of Commerce, University Of Delhi

Paper Received: 14th May, 2021; **Paper Accepted:** 15th June, 2021;
Paper Published: 08th July, 2021

DOI: http://doi.org/10.37648/ijrmst.v11i02.001

How to cite the article:

Prachi Juneja, Employability of Selected Datamining Algorithms in the Early Prediction of Coronary Diseases, IJRMST, July-December 2021, Vol 12, 1-15, DOI: http://doi.org/10.37648/ijrmst.v11i02.001

ABSTRACT

The objective of our work is to take apart unique data mining methods and procedures in the healthcare system that can use an assumption for coronary disease structure and their impact investigation. A coronary disease prediction model, which executes the data mining method, can help the therapeutic experts perceive the coronary sickness status subject to the patient's clinical data. Data mining description techniques for the great fundamental initiative in human system are specifically Decision trees, Naive Bayes, Neural Networks and Support Vector Machines. Hybridizing or merging any of these calculations makes decisions snappier and assigned dynamically. Information mining is a notable new improvement for extracting hypermetropic and critical information from enormous data sets to build significant and novel encounters. Using impelled data mining systems to extract essential information has been considered a fanatic method to improve human management organization's quality and precision while trimming down the social protection cost and execution time. Using this technique can expect the early detection of coronary disease. Using more information properties, for instance, could develop controllable and natural danger factors, progressively detailed results. Can also broaden this strategy. It can use an extensive part of data properties. Other data mining strategies use for forecasts, such as clustering, time series plan, and association rules. The unstructured data open in the human system industry information base can mine using content mining.

INTRODUCTION

The upcoming research has benefitted different researcher in discovering sensible solutions for these issues. Our human organizations are no unique case for this.

Data mining tools have been created for convincing examination of remedial information to help the healthcare experts in improving the treatment purposes. In coronary disease, data mining technique has played out essential work. The particular comprehension between the healthy individuals and the infected heart patient in the already existing therapeutic data is a measurable and unimaginable approach in examining heart-related disease plans to find the hidden remedial information. Coronary disease course of action adds to the unsafe justification for the treatment of patients. To foresee the undeniable nature of coronary disease grounded on the outpouring of the remedial data, the two essential techniques used are Statistics and AI.

Data Mining (DM) is the extended version of information mining in data sets (KDD)1, which extracts specific, uncommon, and possibly essential data.

The differentiation between the mining of data and discovering data is that the latter is the collection of different intelligent computations to divide plans from the data. At the same time, information extraction is the completed system that finds data from metadata. A complete objective is to process high level information from low-level data.

LITERATURE SURVEY

As per Ordonez [1], can predict coronary disorder for certain crucial assets are taken from the patient and in their work have presented a framework that incorporates the qualities of a person dependent on absolutely 13 fundamental labels like sex, pulse, cholesterol and others to anticipate the probability of a patient getting affected by the coronary disorder. They have added two different qualities, for example, fat and smoking conduct and expanded the exploration dataset. For instance, Decision Tree, Naive Bayes, and Neural Network use the information mining classification algorithm to make forecasts, and the outcomes are broken down on the Heart disease data set.

Yılmaz, [2] has proposed a strategy that uses the least squares support vector machine (LS-SVM) using a twofold decision tree to characterize cardiotocograph to discover the patient condition.

Duff, et al. [3] have accomplished examination work including 500 and 33 patients who had experienced heart failure and coordinated them into breaking down coronary illness probabilities. They performed traditional measurable investigation and information mining research utilizing generally Bayesian organizations.

Frawley, et al. [4] have played out a work on the forecast of endurance of coronary illness (CHD), which is a difficult exploration issue for the clinical society. They additionally utilized 10-crease cross-approval strategies to decide the unbiased gauge of the three expectation models for execution examination purposes.

Lee et al. proposed a novel philosophy to extend and consider the multi-parametric component and straight and nonlinear highlights of Heart Rate Variability in diagnosing cardiovascular illness. They have completed different analyses on straight and non-direct highlights to gauge a few classifiers, e.g., Bayesian classifiers, CMAR, C4.5 and SVM. Given their investigations, SVM beat different classifiers.

Noh et al. proposed an arrangement technique; an affiliated classifier developed dependent on the productive FP-development strategy. Since the volume of examples can be different and enormous, they offered a standard to quantify the union and, thus, permit an extreme decision of pruning designs in the model producing measure.

Parthiban, et al. [7] have proposed another work in which coronary disease recognizes and anticipated utilizing the proposed Coactive Neuro-Fuzzy Inference System (CANDIS). Their model works dependent on the aggregate idea of neural organization versatile abilities and dependent on the hereditary prediction and fuzzy logic to analyze the event of the infection. Assessed the presentation of the proposed CANFIS model as far as preparing exhibitions and order exactnesses. At last, their outcomes show that the proposed CANFIS model has an incredible viewpoint in anticipating coronary illness.

Singh, et al. [8] have managed a job utilizing one parcel bunching calculation (K-Means) and one various levelled grouping calculation (agglomerative). K-means algorithm has higher viability and adaptability and merges very fast when created with enormous informational indexes. Different clustering levels develop a progressive system of groups by mixing two more simple collections into more simple ones or parting a bigger group into more modest ones. Utilizing the WEKA information mining tool, they have determined the presentation of k-means and advanced clustering algorithm dependent on precision and execution time. Master et al. [9] have proposed the computational model dependent on a multi-facet perceptron with three layers is utilized to develop a choice emotionally supportive network for the finding of five significant heart disease. The proposed option emotionally supportive network prepared to use a backpropagation calculation intensified with the energy term, the versatile learning rate and the failing to remember mechanics.

Palaniappan et al. [10] have done research work. They have fabricated a model known as Intelligent Heart Disease Prediction System (IHDPS) by utilizing a few information mining strategies like Decision Trees, Naïve Bayes and Neural Network.

Shanta Kumar, et al. [11] have accomplished examination work in which the intelligent and successful respiratory collapse expectation framework is created utilizing Multi-Layer Perceptron with

Back-Propagation. Similarly, the recurrence examples of coronary disease mines with the MAFIA calculation dependent on the information separated.

Yanwei et al. [12] have constructed an arrangement technique dependent on the beginning of multi-parametric highlights by surveying HRV (Heart Rate Variability) from ECG, and the information is pre-handled. A coronary illness expectation model is assembled that arranges the coronary disease of a patient.

METHODOLOGY

To carry out experiment and performances, we use weka as the information mining tool. Weka (Waikato Environment for Knowledge Analysis) is an information mining device written in java created at Waikato. WEKA is an ideal information-mining tool for the clients to characterize the precision dependent on datasets by applying diverse algorithmic methodologies and thought about bioinformatics. Voyager, Experimenter and Knowledge stream are the interfaces accessible in WEKA that we have utilized. In this paper, we have used these information mining methods to foresee the survivability of dengue infection through the grouping of various calculations exactness.

It has four applications:

A. Pioneer: The explorer interface has a few boards like preprocess, order, group, partner, unique characteristic, and Envision. Yet, in this interface, our principle centre is around the Classification Panel.

B. Experimenter: This interface gives an office the precise examination of various calculations dependent on given datasets.

Every calculation runs multiple times, and afterwards, the precision detailed.

C. Information Flow: It is an option in contrast to the Explorer interface. The solitary distinction among this and others is that here client chooses the Weka segment from the toolbar and associates them to make a design for running the calculations.

D. Straightforward CLI: Simple CLI implies order line interface. The client performs tasks through an order line interface by offering directions to the working framework. This interface is less mainstream when contrasted with the other three.

MEDICAL SERVICES DATA MINING

Information digging holds tremendous potential for the therapeutic administration's industry to set up prosperity structures to effectively use data

and assessment for choosing inefficient perspectives and the prescribed procedures that improve the mind and decrease costs. Today, most crisis centers utilize clinical facility information structures to administer large and voluminous patient data proportions. There is a wealth of covered data in this data that is hidden generally. Using data mining2–4 can be changed into supportive information that can empower social protection experts to settle on keen clinical decisions. A bit of the assumption based data mining frameworks are as follows:

Classifiers Based on Bayesian

Applying Bayesian classifiers, the system can track down the personal data identified with sicknesses from chronicled records of patients with coronary disease. Bayesian classifiers anticipate the class investment probabilities to such an extent that the likelihood of a given model has a spot with a particular class. Bayesian classifier relies upon Bayes' theory. We can use the Bayes theory to choose whether a proposed finding is correct, given the insight. A clear probabilistic, the naive Bayes classifier is used for plan subject to which depends on ob Bayes' speculation as shown by an unsuspecting Bayesian classifier the occasion (or non-event of a particular component of a class is clear as independent to the nearest (or non-appearance) of some other element. When the data sources' factor is high, and the doubtless compelling result is typical, the Naïve Bayes Classifier technique5–7 is proper. The naive Bayes model perceives the actual properties and features of patients encountering coronary disease. For every information, it gives the likelihood of value for the anticipated condition. Figure 1 shows how naïve bayes applies in persistent data

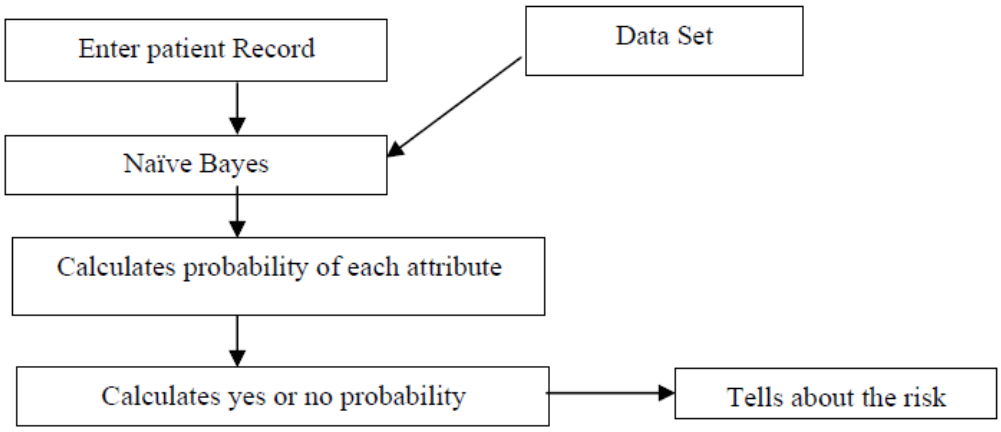


Figure 1. Implementation of Naïve Bayes algorithm on patient data.

Decision Tree

The experts with significant accomplishment have inspected the presentation of the Decision Tree technique8,9 in treating coronary disease. A decision tree is a tree-like design involving internal centres, branches, and leaf centres. Each chapter implies a quality worth. Each inside centre point showed a test on a property used, and a leaf centre point addresses the regular classes or class appointments. The collection starts from the root centre point and explores the tree, subject to the reasonable property assessment. The method incorporates data distributing, data copy, decision tree class decision, and the sales of diminishing an issue slicing to separate decision trees. Game plan procedures are requested as overseen and independent approaches. The managed collection systems contain chi association and entropy, while the independent strategies join undefined width and unclear repeat. The data isolating incorporates testing with or without projecting a voting form. Three Decision Tree types are attempted: Gini Index, Information Gain, and Gain Ratio. Finally, diminished bungle slicing is helpful to give progressively shut decision rules. Figure 2 Shows the Implementation of ID3 estimation on open-minded data.

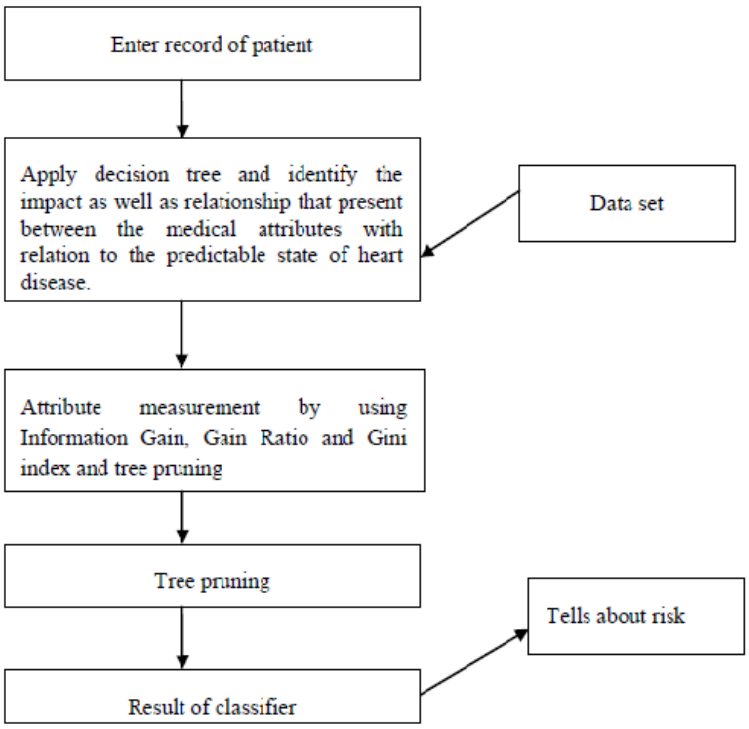


Figure 2. Applying ID3 on patients records

Neural Network

In even-minded applications, neural frameworks can't make significantly accurate results. The neural framework is set up with Heart Diseases information base using feed-forward neural framework model10, variable learning rate and backpropagation learning computation with power. The arrangement of the model is according to the accompanying: It starts with the commitment of clinical data and advances to make ANN evaluation. In the wake of setting up a model, it can convey the figure results. The computational steps of neural framework estimation start with the request for clinical data into equal parts randomly. One is used for testing, and the other is used for training. Total weight is allotted to every segment randomly. The decided mix-ups are used to change the greatness of taking everything into account. Every part's current influence is found when the mix-ups meet with the end conditions. The technique is repeated a few times. To fabricate the planning models, we can sort out the performance results from the testing data. Figure 3 Shows the Implementation of neural framework estimation on clinical data.

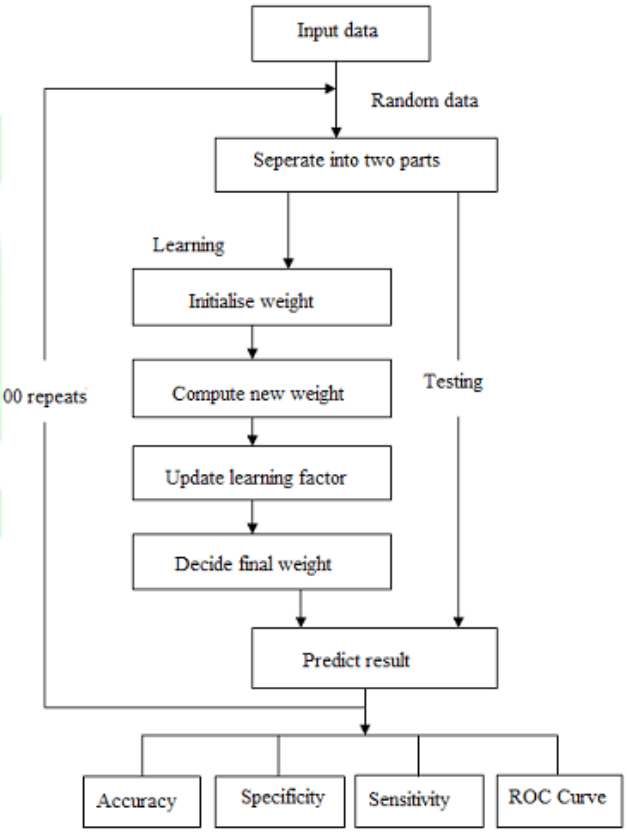


Figure 3. Applying neural network on clinical data

Hybrid

Coronary disease prediction3–5 is one of the enormous troubles in the human administration's industry. Awakened by the in general extending mortality of coronary ailment patients, researchers are using different data mining frameworks to discover coronary sickness. Each strategy has its advantages and negative imprints. Each estimation used by each methodology contains certain limits which are valuable to dissect coronary illness. For an ideal examination of coronary disease, the yields of each computation are merged and considered. Here the mix of output is regarded as "Hybridization". Applying creamer data mining systems can show promising results in the examination of coronary disease. Figure 4 shows the Proposed Approach of the Hybrid Data Mining Technique. Table 1 shows the Data mining instruments and strategies used for coronary ailment assumption with accuracy. Table 2 shows the information mining frameworks used for the examination of different sicknesses.

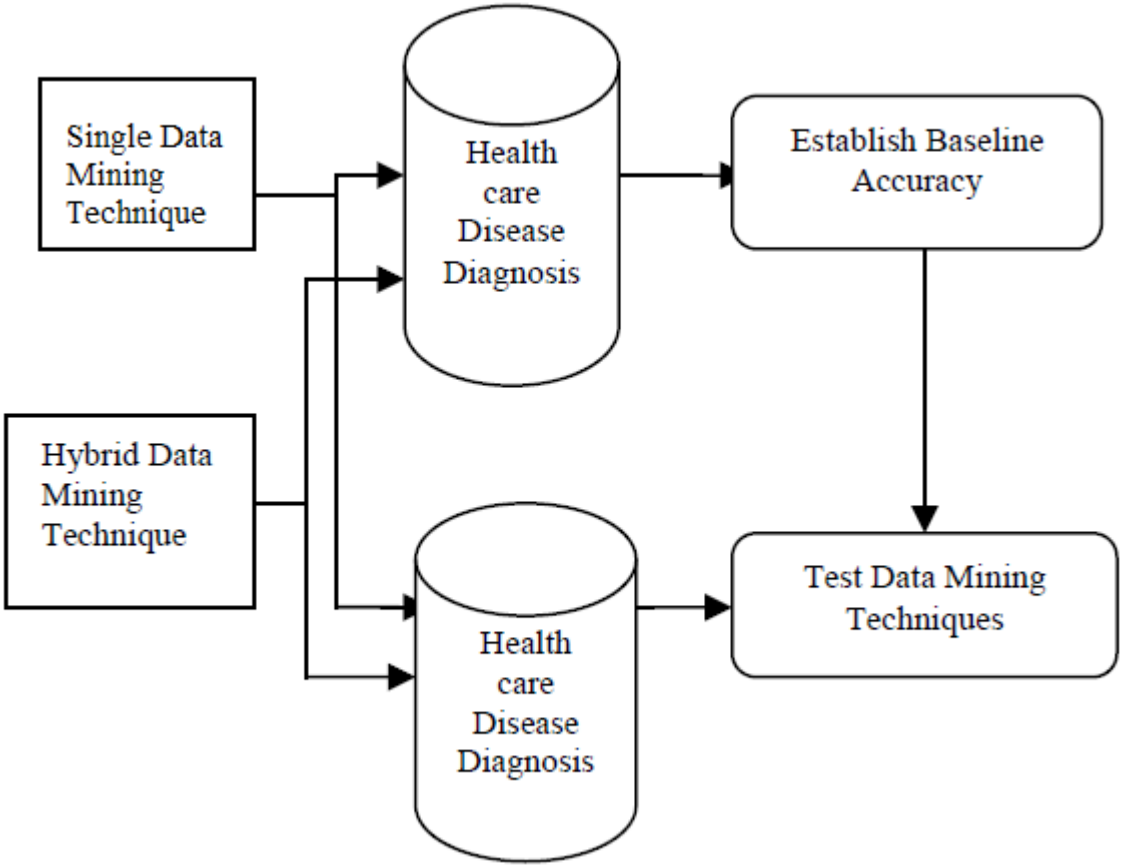


Figure 4. Proposed approach of hybrid data mining technique.

Table 1. Data mining tools and techniques used for heart disease prediction with accuracy

Author	Technique used	Tool used	Accuracy (%)
Abhishek et al. (2013)	J48	Weka 3.6.4	95.56
	Naïve Bayes		92.42
Chaitrali et al. (2012)	Neural Network	Weka 3.6.6	100
Monali et al.	C4.5	Weka	
	Naïve Bayes		
Nidhi et al. (2012)	Naïve Bayes	Weka 3.6.6	100
	Decision Trees	Tanagra	52.33
		Weka 3.6.0	89
	Neural Network	.Net	99.2
Rashedur et al. (2013)	Neural Network	Weka	79.19
	Fuzzy Logic Decision Tree	Tanagra Matlab	83.85
Resul et al. (2009)	Neural Network	SAS base software	89.01

Table 2. Data mining techniques used for diagnosis of different diseases

Author	Year	Disease	Technique
Jaimini Majali et al.	2015	Cancer	Fp-Growth algorithm, ID3 Decision tree
Aiswarya Iyer et al.	2015	Diabetes	Decision Tree and Naïve Bayes
S. Dhamodharan	2014	Liver disease	Naïve Bayesian, Ft Tree
Girija D. K	2013	Fibroid	C4,5, ID3, Naïve Bayes
M. Akhil et al.	2012	Pima Indian diabetes Breast Cancer Heart disease	Associative Classification and Genetic Algorithm
Mohammad et al.	2012	Breast Cancer	C4.5, C5.0
Humar et al.	2008	Diabetes Heart disease	Classification, Backpropagation, Fuzzy Neural Network
Marcel et al.	2007	Carcinoid heart disease	Bayesian Classification

EXPERIMENTAL RESULTS

In this research, the precision of three information mining strategies is analysed. The objective is to have high exactness, other than high accuracy and review measurements. Although these analyses are utilized regularly in data recovery, we have considered them as they are identified with the other existing measures like particularity and affectability. Can get these measurements from the disarray can effortlessly change over the grid to true positive (TP) and false positive (FP) measurements.

A. Confusion Matrix for three classification methods

	a	b
a	67	25
b	17	100

B. *Confusion Matrix for ANN*

	a	b
a	64	28
b	21	96

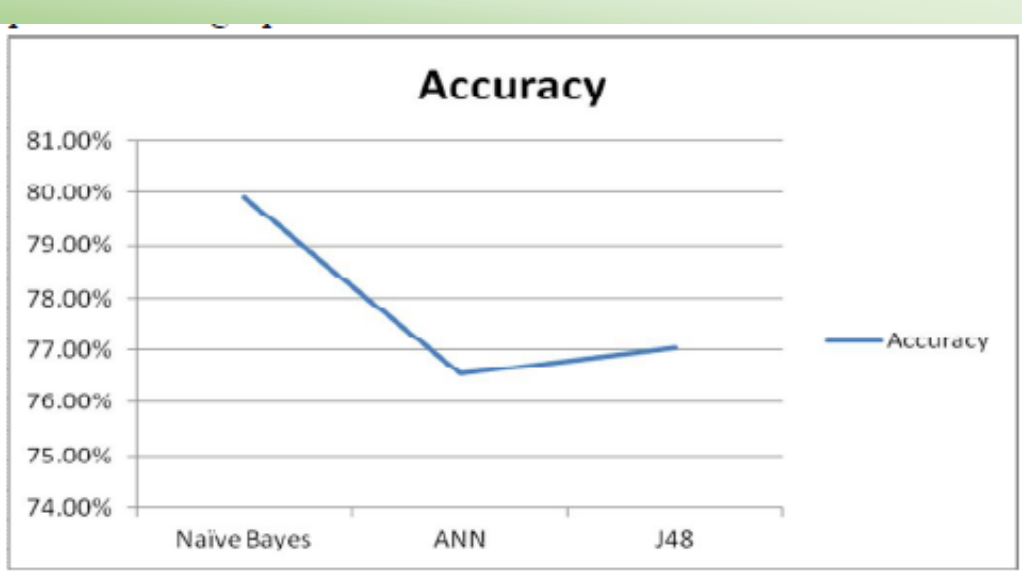
C. *Confusion Matrix for J48*

	a	b
a	66	26
b	22	95

Table Shows accuracy for three different classifications

Classification	Accuracy	Time Taken
Naïve Bayes	79.9043%	0.01 Seconds
ANN	76.555 %	1.55 Seconds
J48	77.0335%	0.01 Seconds

The Accuracy is of each method is plotted on a graph as below:



CONCLUSION

The algorithm's accuracy in every strategy can be improved by hybridizing or consolidating analyses to an individual estimate that may not be exact for feebly grouped informational collections and is relied upon to settle on speedier and more precise choices.

REFERENCES

[1] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542-81554.

[2] Bashir, S., Khan, Z. S., Khan, F. H., Anjum, A., & Bashir, K. (2019, January). Improving heart disease prediction using feature selection approaches. In 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (pp. 619-623). IEEE.

[3] Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics, 36, 82-93.

[4] Ali, L., Rahman, A., Khan, A., Zhou, M., Javeed, A., & Khan, J. A. (2019). An Automated Diagnostic System for Heart Disease Prediction Based on ${\chi^{2}} $ Statistical Model and Optimally Configured Deep Neural Network. IEEE Access, 7, 34938-34945.

[5] Maji, S., & Arora, S. (2019). Decision tree algorithms for prediction of heart disease. In Information and Communication Technology for Competitive Strategies (pp. 447-454). Springer, Singapore.

[6] Khourdifi, Y., & Bahaj, M. (2019). Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. Int. J. Intell. Eng. Syst., 12(1), 242-252.

[7] Latha, C. B. C., & Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked, 16, 100203.

[8] Ayon, S. I., Islam, M. M., & Hossain, M. R. (2020). Coronary artery heart disease prediction: a comparative study of computational intelligence techniques. IETE Journal of Research, 1-20.

[9] Burse, K., Kirar, V. P. S., Burse, A., & Burse, R. (2019). Various preprocessing methods for neural network-based heart disease prediction. In Smart innovations in communication and computational sciences (pp. 55-65). Springer, Singapore

[10] Tarawneh, M., & Embarak, O. (2019, February). Hybrid approach for heart disease prediction using data mining techniques. In International Conference on Emerging Internetworking, Data & Web Technologies (pp. 447-454). Springer, Cham.

[11] Beunza, J. J., Puertas, E., García-Ovejero, E., Villalba, G., Condes, E., Koleva, G., ... & Landecho, M. F. (2019). Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease). Journal of biomedical informatics, 97, 103257.

[12] Gonsalves, A. H., Thabtah, F., Mohammad, R. M. A., & Singh, G. (2019, July). Prediction of coronary heart disease using machine learning: an experimental analysis. In Proceedings of the 2019 3rd International Conference on Deep Learning Technologies (pp. 51-56).

[13] Nalluri, S., Saraswathi, R. V., Ramasubbareddy, S., Govinda, K., & Swetha, E. (2020). Chronic Heart Disease Prediction Using Data Mining Techniques. In Data Engineering and Communication Technology (pp. 903-912). Springer, Singapore.

[14] Gokulnath, C. B., & Shantharajah, S. P. (2019). An optimized feature selection based on genetic approach and support vector machine for heart disease. Cluster Computing, 22(6), 14777-14787.

[15] Enriko, I. K. A. (2019, June). Comparative study of heart disease diagnosis using top ten data mining classification algorithms. In Proceedings of the 5th International Conference on Frontiers of Educational Technologies (pp. 159-164).

[16] Singh, A., & Kumar, R. (2020, February). Heart Disease Prediction Using Machine Learning Algorithms. In 2020 International Conference on Electrical and Electronics Engineering (ICE3) (pp. 452-457). IEEE.

[17] Barik, S., Mohanty, S., Rout, D., Mohanty, S., Patra, A. K., & Mishra, A. K. (2020). Heart Disease Prediction Using Machine Learning Techniques. In Advances in Electrical Control and Signal Systems (pp. 879-888). Springer, Singapore.

[18] A. Powar, S. Shilvant, V. Pawar, V. Parab, P. Shetgaonkar and S. Aswale, "Data Mining & Artificial Intelligence Techniques for Prediction of Heart Disorders: A Survey," 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India, 2019, pp. 1-7, doi: 10.1109/ViTECoN.2019.8899547.

[19] Kumari, M., & Godara, S. (2011). Comparative study of data mining classification methods in cardiovascular disease prediction 1.

[20] Dangare, C. S., & Apte, S. S. (2012). Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications, 47(10), 44-48

[21] Ashish C, Lakhan A, Sahil A and Prof Y.K.Sharma, P.(2016). Heart Disease Prediction Using Data Mining Techniques. International Journal Of Research in Advent Technology.

[22] Das, Resul, Turkoglu, Ibrahim, et al.: Effective diagnosis of heartdisease through neural networks ensembles. J. Expert Syst. Appl.36, 7675–7680 (2009)

[23] Das, Resul, Turkoglu, Ibrahim, et al.: Diagnosis of valvular heart disease through neural networks ensembles. J. Comput. Methods Progr. Biomed. 93, 185–191 (2009)

[24] Gokulnath, C., Priyan, M. K., Balan, E. V., Prabha, K. R., Jeyanthi, R.: Preservation of privacy in data mining by using PCA based perturbation technique. In: Smart Technologies and Man- agement for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015 International Conference on (pp.202–206). IEEE (2015)

[25] Babaoglu, et al.: Assessment of exercise stress testing with arti- ficial neural network in determining coronary artery disease and predicting lesion localization. J. Expert Syst. Appl. 36, 2562–2566 (2009)

[26] Rajeswari, K., et al.: Feature selection in ischemic heart disease identification using feed forward neural networks. Int. Symp. Robot. Intell. Sens. 41, 1818–1823 (2012)

[27] Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., & Steinberg, D. Top 10 algorithms in data mining. Knowledge and Information Systems, 14(1), 1-37 (2008).

[28] Tilve, A., Nayak, S., Vernekar, S., Turi, D., Shetgaonkar, P. R., & Aswale, S. (2020, February). Pneumonia Detection Using Deep Learning Approaches. In 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) (pp. 1-8). IEEE.