A Review of Role of Machine Learning Models in Coronary Heart Disease Detection Accuracy

Tarun Kumar Agarwal[1], *Hemant Sharma*[2], *Challa Madhavi Latha*[3], *and Sitaram Gupta*[4]

Abstract

According to the World Health Organization, heart disease is the most widespread disease in the world, affecting over a billion people. Generally, the lifestyles of people are occasionally plagued by stress, worry, and sadness, among other things. The early detection of this condition is tough, and it is a difficult task in medical science. The goal of this research is to better understand the detection accuracy of particular machine learning models (MLMs), as well as their limitations and categorization strategies. Many researchers used classification techniques such as Naive Bays (NB), decision trees (DT), Cooperative Neural-Network Ensembles (CNNEs), logistic regression (LR), Support Vector Machine (SVM), Least Square Twin Support Vector Machine (LS-SVM), k-Nearest Neighbor (KNN), Bays Net (BN), Artificial Neural Network (ANN), and Multi-Layer Perception (MLP) (MLP). In total, the dataset contains more than 50 features attributes. To boost accuracy, the study uses different feature selection approaches to identify the most appropriate features for detecting the disease. The present study achieved a maximum classification accuracy of 96.29%, and there is a need to improve accuracy in the shortest period possible by developing single MLMs for detecting and selecting specific features. Many studies employ hybrid approaches to improve the accuracy of percentages by layering two or more classification algorithms (based on specified symptoms and traits of a human being). It is not always more efficient and time-consuming. As a result, flexible MLMs with feature selection and reduction strategies are required. Further, the current research focuses on boosting accuracy and includes future viewpoints or uses of research as well.

Keywords : **Artificial Neural Network, Cooperative Neural-Network Ensembles, K-Nearest Neighbor, Least Square Twin Support Vector Machine, Multi-Layer Perception, Naive Bays, Support Vector Machine**

I. INTRODUCTION

In medical science, a healthy heart is the sign of a healthy life. A large number of people are suffering from coronary heart disease as per the medical industry data and it is a common disease because of which several people die. So, the detection of this disease is a priority for healthcare. There are four chambers in the heart, two atriums and two ventricles. In general, the most common heart disease is Coronary heart disease (CHD) in which plaque deposits block the coronary blood vessels and reduce the supply of blood and O_2 to the heart [30, 1, 16, 11, 9, 10]. As per healthcare, heart disease is caused by sedentary lifestyle, food habits, family history, stress, diabetes, lack of

Manuscript Received : October 5, 2021 ; Revised : December 5, 2021; Accepted : December 21, 2021. Date of Publication : February 5, 2022.

T. K.Agarwal[1] is *Assistant Professor & Research Scholar* with Department of CSE, Vivekananda Global University, Jaipur - 303 012, Rajasthan. Email : tarun.agarwal@vgu.ac.in ; ORCID iD : https://orcid.org/0000-0001-7805-3282
H. Sharma[2] is *Assistant Professor* with Department of CSE, Vivekananda Global University, Jaipur - 303 012, Rajasthan. Email : hem.s1209@gmail.com ; ORCID iD : https://orcid.org/0000-0003-2571-4854
C. M. Latha[3] is *Assistant Professor* with Department of CSE, CMR College of Engineering & Technology, Medchal, Hyderabad, Telangana - 501 401. Email : drchmadhavilatha@cmrcet.ac.in ; ORCID iD : https://orcid.org/0000-0003-0861-9513
S. Gupta[4] is *HOD,* Department of Computer Science at Vivekananda Global University, Jaipur, Rajasthan - 303 012. Email : sitaram.gupta@vgu.ac.in ; ORCID iD : https://orcid.org/0000-0003-4446-7633

DOI: https://doi.org/10.17010/ijcs/2022/v7/i1/168955

exercise, excess weight, age, cholesterol, high blood pressure, smoking, drug abuse, fast blood sugar, alcohol, etc. [24, 28, 2, 10, 23].

↳ ***Naive Bays (NB)*** is a simplified version of the Bayes formula to decide which class a feature belongs to. The posterior probability of each feature, the feature values are given; we assign instances class with the highest probability. According to the concept of the Bayesian theorem,

$$p(x|y) = p(x) * p(y/x)/p(y)$$
$$\text{where, } p(y|x) = p(\text{variable } x \cap y)/p(x) \quad (1)$$

$p(\text{variable } x)$= the probability of x happening

$p(y)$ = the probability of y happening

$p(x|y)$ = the probability of x given y

$p(y|x)$ = the probability of y given x

$p(x \cap y)$ = the probability of both x and y happening

Simple and effective for the prediction process: Based on eq. (1), the Bayesian classifier calculates each class that belongs to the conditional probability of an instance that can work with one class mode and/or multiclass mode. Fewer statistics are needed for training to construct the model. Better performance is required from researchers to hold sovereignty upto the possible levels and attain max results with categorical input variables supposing simple distribution for the numerical variable. Assign a 0 (zero) probability for the real data if data is not present at the training time of the model, that is, not present within the training dataset. This issue can be cracked by using the Laplace estimation technique [3, 7, 10, 32].

Cooperative Neural-Network Ensembles (CNNEs), the proposed methodology emphasizes disease diagnosis accuracy and diversity among individual Neural Network (NNs). The approach defines and calculates the number of the hidden vertex in individual NNs by a constructive approach and uses incremental training based on negative correlation for training individual NNs over training epochs. An experimental result shows that this methodology produces good generalization ability over individual NNs [4, 5, 6, 24, 28].

Model LSTSVM (Least Square Twin Support Vector Machine) is a machine learning method that achieved the highest accuracy to diagnose heart diseases with only eleven features set. However, authors of this approach have not evaluated feature selection mechanisms with other accessible techniques. On the other hand, techniques have not been analyzed with different data sets. Hence, there is a chance that this technique may produce low accuracy with other data sets, so it is not a good solution for the classification of heart disease [30, 23, 4, 25].

K-Nearest Neighbor (KNN) based approaches are the most frequently and significantly used data mining techniques in classification issues or problems. It is showing good performance in classification problems with various datasets but the main drawback of KNN classifiers is the high memory requirement needed to keep the whole sample. If the sample is large, the response time on a sequential system is also large. Despite the need for memory issues, the use of a simple KNN algorithm is not a good choice and requires an optimization process [29, 15, 17,11].

The prediction of the degree of angiographic CHD by Artificial Neural Network (ANN) and subsequently, its performance is evaluated using a coronary heart disease database produced from Cleveland Clinic Foundation UCI Library Database with the entire numeric-valued attributes. About 88 cases of certain kinds of aged angiographic coronary heart disease subjects with 13 attributes already tested in ANN were considered and the study showed that ANN-based prognosis of CHD improved the diagnostic accuracy prediction result to 95.5% which is maximum or highest with recent works [30, 14, 20, 26, 5, 7, 27, 22].

Multilayer perception (MLP) is a feed-forward neural network, a form of artificial neural network (ANN) which maps the datasets of applying input data into a set of relative outputs. The term feedforward means that data flows in a single direction from the input to the output layer. It uses a backpropagation learning procedure for the training purpose of the designed network and is more capable of resolving an issue that is not linearly separable. Therefore, it is widely used for pattern prediction, detection classification, and approximation. In simple terms, a directed graph MLP consists of multiple layers of nodes, with every layer connected to the next one. Except for the input nodes, a nonlinear activation function, each node is a neuron. Backpropagation for training the network known as MLP utilizes a supervised learning technique. MLP can differentiate data that cannot be linearly divided [19, 13, 28, 17, 8].

This technique has the following advantages:

- The neural network is proficient in managing extensive troubles.
- Can outperform even in complicated domains.
- Efficiently handles both categorical and continuous data types.
- Designed to be parallelized.

Decision Tree (DT) is related to the information entropy principle and it is the most used classification technique. It is used to find a significant feature by making and pruning trees. This information is developed in every node to construct the decision tree. All possible features are analyzed from the head node and maximum information features node is considered for evaluation and less information node is eliminated [30, 12, 7, 15, 16].

Gaps in published research: The researchers had presented many techniques and methodologies on techniques for diagnosing heart disease. All the researchers had shown improvement. The gaps observed in total were as follows: most of the solution approaches in the literature review are based on the working concept of traditional prediction algorithms that use a hybrid approach to predict diseases with higher accuracy. Some approaches show only the theoretical discussion rather than the proof. Numerous algorithms presented by different investigators use different concepts but still no common approach is available which can outperform or maintain its accuracy with different datasets [9, 28, 12, 14].

II. EVALUATION PARAMETERS USED TO ANALYZE DETECTION PERFORMANCE OF A PARTICULAR MODEL

Confusion Matrix

Performance analysis as per the proposed model is a very popular methodology known as confusion matrix that has been used in this work. It is also known by the name of coincidence matrix, classification matrix, or contingency table. Table I shows a two-class classification problem.

The left-most columns indicate data predicted by the models. The diagonal data indicates correct predictions. In the case of a positive instance it is classified as positive, it is treated and counted as a true positive (TP); if it is classified as negative, it is counted as a false negative (FN). In the case of a negative instance, it is classified as negative, it is treated and counted as a true negative (TN); In the case of classification as positive, it is treated and counted as a false positive (FP). These terms are helpful when analyzing a classifier's ability.

The prediction performance of the proposed scheme is calculated based on accuracy, measured by the following equation.

$$\textbf{Accuracy(\%)} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

The actual percentages predicted by the classification model are represented by False Negative (FN), False Positive (FP), True Positive (TP), True Negative (TN). A classification algorithm may wrongly classify some instances because of the biased nature of the dataset but this may appear to be accurate. As a result, to compare accuracy, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) rates have been calculated to find out the main accuracy of a simulated classification algorithm.

(1) True Positive (TP) : Rate of correctly predicted data against its class type.

(2) True Negative (TN) : A situation, data examined has taken its correct position, and no alarm is raised.

(3) False Positive (FP) : A state that data has not been associated with its relative class.

(4) False Negative (FN) : A state when a class data is predicted as another class type data [28, 30, 31, 26, 17, 20, 21, 22].

As shown in Fig. 1, the process flow diagram provides

TABLE I.
CONFUSION MATRIX FOR TWO-CLASS CLASSIFICATION PROBLEM

		Predicted Class	
		Positive	**Negative**
Actual reclass	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

the details of experimental work carried out by researchers in the past. First, we get the data (optimum data sets) from the UCI library which provides the open-source datasets for both testing and training data. After this, we select the optimal feature selection or reduction techniques for data processing and then create testing and training datasets and apply machine learning models (it may be hybrid or single approach) as per accuracy of results and finally, analyze the results with past information or data.

III. DISCUSSION

In this section, we discuss Table II, a comparison of various MLMs, Feature Selection, and Feature Reduction Techniques with results as well as limitations/weaknesses. The coronary heart disease dataset is considered for further study and knowledge purposes, and various machine learning models are used to detect heart disease with Feature selection and deduction techniques. The target of this review is to compare the accuracy in the percentage of different classification MLMs and present the topmost one. Based on all the results, different techniques are performed better according to the researcher (with or without validation). However, every technique has intrinsic power to outperform other techniques depending on the situation [15, 17, 18].

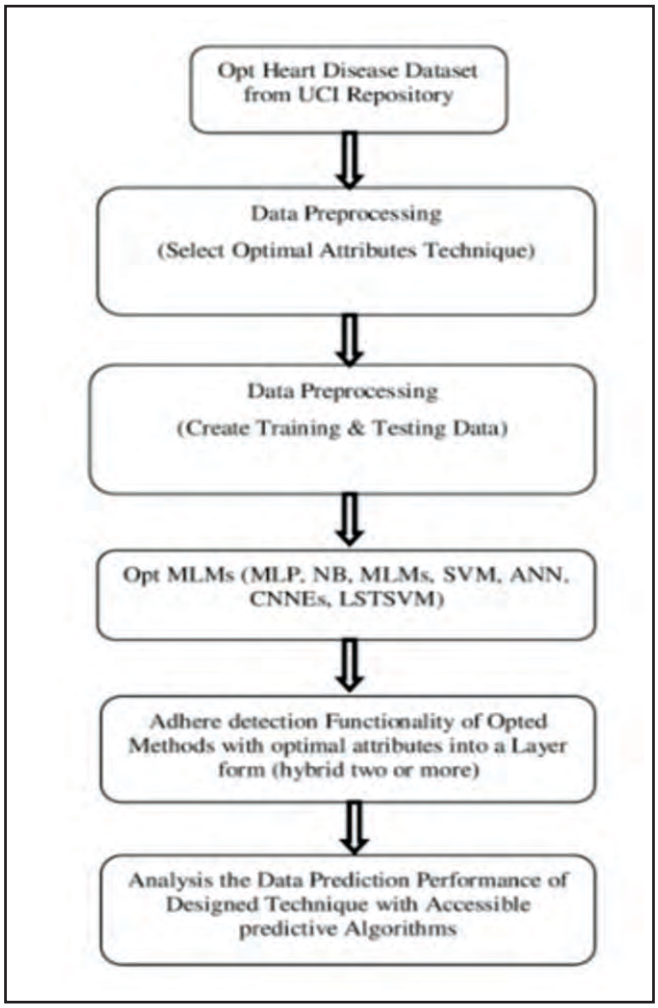

Fig. 1. Workflow adopted by researchers

In this review process, the performance of base MLMs and Features Select/Reduction Techniques (Table I) is done. With SVM and MLP hybrid binary dragonfly algorithm (BDA), mutual information (a filter method), accuracy of 96.29% was achieved and mutual information is not effective for prediction because sometimes mutual information is not feasible to improve accuracy.

On the other hand, the performance of Hybrid MLMs (DT, NB & RF) shows that probability of predicting heart disease was 95% and sometimes, a diagnostic test is positive and the technique is accurate in preventing and making an accurate prediction of heart disease [1, 2, 5, 28, 16, 29, 8, 9, 10]. Many MLMs with feature selection and reduction used by a researcher are not accurate. Sometimes, they hide the precise features related to disease at the time of the process and show less than 95% accuracy after removal or elimination of various disease related features which are not effective. Moreover, the researchers are focusing on improving accuracy results, but sometimes 100% accuracy is not considered in medical industry. In medical science, every feature in the duration of the entire disease treatment is noticed and observed but NB, SVM, DT, and MLP are very effective models for this disease in terms of detection.

IV. RESULTS

Table I shows that Liu et al. [29] used K-NN, NB, and ANN MLMs with feature space mapping (FSM), as well as separability split value (SSV); both are feature selection techniques, and ReliefF and Rough Set (RFRS) methods remove superfluous and redundant features more effectively. Accuracy of 92.59% was achieved which is low but effective for future perspectives. The weakness is in both the weight threshold and the number of the nearest neighbors in the ReliefF algorithm [29].

TABLE II.
COMPARISON OF VARIOUS MLMS

Reference	MLMs	Feature selection techniques adopted by the researcher in the past	Feature reduction techniques adopted by the researcher in the past	Accuracy (%)	Limitation/Weaknesses
Liu et al. [29]	K-NN+NB+ANN	Separability split value (SSV) and Feature space mapping (FSM).	ReliefF and Rough Set (RFRS) method remove superfluous and redundant features more effectively.	Accuracy of 92.59% was achieved.	Both weight threshold and number of the nearest neighbors in the ReliefF algorithm.
Agrawal [28]	NB+MLP	Correlation-based feature selection (CFS).	Best First Search (BFS)	Accuracy of **95.19%** was achieved.	High rate of false prediction and highly dependent on the performance of the first technique.
Tomar and Agarwal [32]	LSTSVM	*F*-score feature selection technique calculates the weight of every feature and higher weight is chosen as a priority.	–	Accuracy of **94.63%** was achieved.	Dependent on features weight and not effective (only 11 features set).
Saeed and Al-Ta'I [20]	SVM+MLP	Hybrid binary dragonfly algorithm (BDA).	Mutual information (a filter method).	Accuracy of **96.29%** was achieved.	Mutual information is not effective for the prediction
Wenxin [30]	SVM+ANN+DT	Standard Scalar method (SSM), and ensure the numerical frequency of every feature is divided into 0 to 1.	–	Accuracy of **95%** was achieved	Useful for doctors but not effective and more time-consuming.
Das, Naik and Behera [7]	ANN+**Neuro-Fuzzy**		Neuro-Fuzzy and Feature Reduction (NF-FR) model used a feature-based class belongingness fuzzification process for all the patterns and expanded based on the number of classes available in the dataset. It is used to filter out the insignificant features.	Accuracy of **95.15%** was achieved.	Statistical analysis is not for real-life problems.
Kamley [24]	**SVM+Logistic Regression(LR)**	Best features : Information gain is used to find the features.		Accuracy of not more than **85%** was achieved.	Take five or six common features only which are fixed.
Lin, Yang, Lin, and Fu [6]	**NN and CNN**	In NN the numbers of hidden layers, with the neuron number greater than 100 in each layer, and CNN the numbers of hidden layers in the fully connected network (FCNet).		Accuracy of not more than **93.81%** was achieved.	It worked only for Cleveland Heart Disease Data Set and compared the prediction accuracy of the two models under different parameters settings.
Grampurohit and Sagarnal [27]	**DT+Random Forest (RF), and NB**	Out of 132, 95 symptoms are considered independent variables closely matched to diseases were selected and optimized but the dependent variable was composed of 41 only.		Accuracy of not more than **95%** was achieved.	Comprehensive comparative study of three MLMs and the performance is analyzed through confusion matrix and accuracy only.
Senthil Kumar Mohan et al.	**RF +** Linear Method (LM)	Hybrid random forest with a linear Model (HRFLM) is used for combining the characteristics.		Accuracy of not more than **88.7%** was achieved.	New feature selection methods can be developed.

Agrawal [28] used NB and MLP with Correlation-based feature selection (CFS). With Best First Search (BFS), accuracy of 95.19% was achieved. The weakness is the high rate of false prediction that is highly dependent on the performance of the first technique and its select minimum features , which is sometimes not considered in the medical industry [28, 20, 26]. It reported higher accuracy with RF than CNN and NN as reported by other authors [6, 21, 9]. Tomar and Agarwal [32] used the LSTSVM F-score feature selection technique which calculates the weight of every feature, and a higher weight is chosen as a priority. Accuracy of 94.63% was achieved and it depends on features weight and is not effective (only 11 features set) [9, 22, 23]. Saeed and Al-Ta'I [20] used the SVM and MLP hybrid binary dragonfly algorithm (BDA). With mutual information (a filter method) accuracy of 96.29% was achieved which is the highest accuracy in terms of prediction and the limitation is that mutual information is not effective for prediction [9, 28, 25].

Wenxin [30] SVM+ANN+DT Standard Scalar method (SSM), ensures the numerical frequency of each feature is divided from 0 to 1. Accuracy of 95% was achieved and it is useful for doctors but is not effective and more time-consuming. It is better than the MLP and NB hybrid approach [24]. Himansu Das et al. used ANN, and Neuro-Fuzzy and applied Neuro-Fuzzy and Feature Reduction (NF-FR) model. They used a feature-based class belongingness fuzzification process for all the patterns and expanded based on the number of classes available in the dataset. It is used to filter out insignificant features. Accuracy of 95.15% was achieved. Statistical analysis is not for real-life problems, because there are several better approaches such as SVM, LR, NN, CNN, and LSTSVM machine learning approaches [24, 6, 9, 19]. Kamley [24] used SVM and Logistic Regression (LR). Best features Information gain is used to find the features. Accuracy of not more than 85% was achieved which is very less and not more effective in the medical industry and it takes five or six common features only which are fixed. Lin et al. [6] used the NN and CNN. In NN the number of hidden layers with the neuron number greater than 100 in each layer, and CNN the numbers of hidden layers in the fully connected network (FCNet). Accuracy of not more than 93.81% was achieved and it worked only for Cleveland Heart Disease Data Set. Comparing the prediction accuracy of the two models under different parameter settings, it is better than the SVM and Logistic Regression (LR) hybrid approach [9, 27]. Sneha and Chetan [27] used the DT, Random Forest (RF), and NB and out of 132, 95 symptoms considered as independent variables that closely matching the diseases were selected and optimized. Symptoms closely related to diseases were selected and optimized. Accuracy of not more than 95% was achieved. A comprehensive comparative study of three MLMs and their performance was analyzed through confusion matrix and accuracy only [9, 26]. In [34] the researchers used the RF and Linear Method (LM) hybrid random forest with a linear Model (HRFLM) to combine the characteristics. Accuracy of not more than 88.7% was achieved. New feature selection methods can be developed. However, 88.7% accuracy can be considered in medical science.

V. CONCLUSION

The review process carried out in this investigation has defined many novel hybrid techniques and approaches to data detection and prediction by incorporating the prediction functionality of MLMs of data mining techniques into a single practice. The designed approach of the present investigation is not simply integrated with Machine learning algorithms for data prediction but also considered selection process attributes at the time of building a technique. For the selection of only relative and result-oriented attributes feature selection technique has been used. In this review process, the result has been analyzed. The results for features selection and reduction have been achieved with the best accuracy of MLMs. Moreover, decision-making on the best suitable path feature selection is required to improve the accuracy of disease in the future. Reducing required data evaluation time may improve the performance of the presented techniques in a new way. Inclusion of more competent techniques of choosing optimal features from elected datasets would improve or increase prediction performance efficiently. The illustration of a single approach with the combination of two or more MLMs by applying layered form approaches rather than hybrid approaches will efficiently lead to fast detection.

AUTHORS' CONTRIBUTION

Tarun Kumar Agarwal made a significant addition to the article's concept or idea. Hemant Sharma was in charge of

the article's analysis. Madhavi Latha wrote the article and critically edited it for intellectual substance. Sitaram Gupta gave his approval to the final version to be published. All members have committed to be responsible for all parts of the work, including ensuring that any questions about the accuracy or integrity of any portion of it are thoroughly examined and resolved.

CONFLICT OF INTEREST

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

FUNDING ACKNOWLEDGEMENT

The authors received no financial support for this research, its authorship or for its publication.

REFERENCES

[1] A. T. Sayad and P. P. Halkarnikar, "Diagnosis of heart disease using neural network approach," *Int. J. Advances Sci. Eng. Technol.,* vol. 2, no. 3, pp. 88–92, Jul. 2014. http://www.iraj.in/journal/journal_file/journal_pdf/6-71-140490825388-92.pdf

[2] A. Taneja, "Heart disease prediction system using data mining techniques," *Oriental J. Comput. Sci. Technol.,* vol. 6, no. 4, pp. 457–466, Dec. 2013.

[3] A. Methaila, P. Kansal, H. Arya, and P. Kumar, "Early heart disease prediction using data mining techniques," in *7th Int. Conf. Comput. Sci., Eng. Inform. Technol. (CCSEIT 2017), Comput. Sci. Inform. Technol.,* pp. 53–59, 2014. [Online]. Available: https://airccj.org/CSCP/vol4/csit42607.pdf

[4] B. A. Jabbar, B. L Deekshatulu, and P. Chandra, "Classification of heart disease using artificial neural network and feature subset selection," *Global J. Comput. and Technol. Neural Artif. Intell.,* vol. 13, no. 3, pp. 5–14, 2013. [Online]. Available: https://core.ac.uk/download/pdf/231159766.pdf

[5] C. S. Prakash, M. Madhu Bala, and A. Rudra, "Data Sci. framework - Heart disease predictions, variant models and visualizations," in *2020 Int. Conf. Comput. Sci., Eng.Appl.,* 2020, pp. 1–4, doi: 10.1109/ICCSEA49143.2020.9132920

[6] C. -H. Lin, P.-K. Yang, Y. -C. Lin and P. -K. Fu, "On machine learning models for heart disease diagnosis," in *2nd IEEE Eurasia Conf. Biomed. Eng., Healthcare Sustainability*, 2020.

[7] H. Das, B. Naik, and H. S. Behera, "A hybrid neuro-fuzzy and feature reduction model for classification," *Advances Fuzzy Syst.,* vol. 2020, Article ID 4152049, pp. 1–15, 2020, doi:10.1155/2020/4152049

[8] D. Vadicherla and S. Sonawane, "Classification of heart disease using SVM and ANN," *Int. J. Res. Comput. Commun. Technol.,* vol. 2, no. 9, pp. 694–701, Sep. 2013.

[9] M. Durairaj and V. Revathi, "Prediction of heart disease using back propagation MLP algorithm," *Int. J. Scientific Technol. Res.,* vol. 4, no. 8, pp. 235–239, Aug. 2015.

[10] H. Murthy and M. Meenakshi, "ANN model to predict coronary heart disease based on risk factors," *Bonfring Int. J. Man-Mach. Interface,* vol. 3, no.2, pp. 13–18, Jun. 2013.

[11] W. H. Hong, J. H Yap, and G. Selvachandran, P. H. Thong, and L. H. Son, "Forecasting mortality rates using hybrid Lee–Carter model, artificial neural network and random forest," *Complex Intell. Syst.,* vol. 7, pp. 163–189, 2021. doi:10.1007/s40747-020-00185-w

[12] J.Banupriya and S. Kiruthika, "Heart disease using data mining algorithm on neural networks and genetic algorithm," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.,* vol. 6, no. 8, pp. 40–42, Aug. 2016.

[13] K. S. Krishnasree and M. R. N. Rao, "Diagnosis of heart disease using neural networks - Comparative study of Bayesian regularization with multiple regression model," *J. Theor. Appl. Inform. Technol.,* vol. 88, no. 3, pp. 638–643, Jun. 2016.

[14] K. Vinay R., K. L. S. Soujanya, and P. Singh, "Disease prediction by using deep learning based on patient treatment history," *Int. J. Recent Technol. Eng.,* vol. 7, no. 6, pp. 1159–1168, Mar. 2019.

[15] C. M. Latha and K. L. S. Soujanya, "Enhancing end-

to-end device security of internet of things using dynamic cryptographic algorithm," *Int. J. Civil Eng. Technol.* vol. 9, no. 9, pp. 408–415, 2018.

[16] L. Verma, S. Srivastava, and P. C. Negi, "An intelligent noninvasive model for coronary artery disease detection," *Complex Intell. Syst.,* 4, pp. 11–18, 2018, doi: 10.1007/s40747-017-0048-6

[17] C. M. Latha and K. L. S. Soujanya, "Secure IoT Framework Through FSIE Approach," in Singh P. K., Veselov G., Vyatkin V., Pljonkin A., Dodero J. M., Kumar Y. (eds), "*Futuristic trends in network and communication technologies,*" FTNCT 2020. *Commun. Comput. Inf. Sci.,* vol. 1395, Springer, Singapore, 2021, doi: 10.1007/978-981-16-1480-4_2

[18] M. G. Tsipouras, T. P. Exarchos, D. I. Fotiadis, A. P. Kotsia, K. V. Vakalis, K. K. Naka, and Lampros K. Michalis, "Automated diagnosis of coronary artery disease based on data mining and fuzzy modeling," *IEEE Trans. Inf. Technol. Biomedicine,* vol. 12, no. 4, pp. 447–458, 2008, doi: 10.1109/TITB.2007.907985

[19] M. A. M. Abushariah, A. A. M. Alqudah, O. Y. Adwan, and R. M. M. Yousef," Autom. heart disease diagnosis system based on Artif. Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Syst. (ANFIS) approaches," *J. Softw. Eng. Appl.,* 7, pp. 1055–1064, 2014.

[20] N. A. Saeed and Z. T. M. Al-Ta'I, "Feature selection using hybrid dragonfly algorithm in a heart disease predication system," *Int. J. Eng. Adv. Technol.,* vol. 8, no. 6, pp. 2862–2867, 2019, doi:10.35940/ijeat.f8786.088619

[21] P. Bajaj and P. Gupta, "Review on heart disease diagnosis based on datamining techniques," *Int. J. Sci. Res.,* vol. 3, no. 5, pp. 1593-1596, May 2014.

[22] R. J. P. Princy, S. Parthasarathy, P. S. H. Jose, A. R. Lakshminarayanan, and S. Jeganathan, "Prediction of cardiac disease using Supervised Machine Learning Algorithms," *Proc. Int. Conf. Intell. Comput. Control Sys.,* May 13–15, 2020.

[23] S. Florence, N. G. B. Amma, G. Annapoorani, and K. Malathi, "Predicting the risk of heart attacks using neural network and decision tree," *Int. J. Innovative Res. Comput. Commun. Eng.,* vol. 2, no. 11, pp. 7025–7030, Nov. 2014.

[24] S. Kamley, "Performance of hybrid ensemble classification techniques for prevalence of heart disease prediction," *Int. J. Innovative Technol. Exploring Eng.,* vol. 8, no. 10, pp. 1875–1882, Aug. 2019, doi:10.35940/ijitee.j9233.0881019

[25] S. S. Sanagala, S. K. Gupta, V. K. Koppula, and M. Agarwal, "A fast and light weight deep convolution neural network model for cancer disease identification in human lung(s)," in *2019 18th IEEE Int. Conf. Mach. Learn. Appl., (ICMLA) 2019,* Boca Raton, FL, USA, Dec. 16–19, 2019, pp. 1382–1387, doi: 10.1109/ICMLA.2019.00225

[26] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," in *IEEE Access,* vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707

[27] S. Grampurohit and C. Sagarnal, "Disease prediction using machine learning algorithms," in *2020 Int. Conf. Emerg. Technol. (INCET),* 2020, pp. 1–7, doi: 10.1109/INCET49848.2020.9154130

[28] T. K. Agrawal, "Neural network & Naïve Bays based hybrid model for heart disease diagnosis," in *12th Biyani Int. Conf.* (BICON-2017), 2017.

[29] X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu, Q. Wang, and Q. Wang, "A hybrid classification system for heart disease diagnosis based on the RFRS method," *Comput. Math. Methods Medicine,* vol. 2017, Art. No. 8272091, pp. 1–11, doi: 10.1155/2017/8272091

[30] X. Wenxin, "Heart disease prediction model based on model ensemble," in *2020 3rd Int. Conf. Artif. Intell. Big Data (ICAIBD),* 2020, pp. 195–199, doi: 10.1109/ICAIBD49809.2020.9137483

[31] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network - Genetic algorithm," *Comput. Methods Programs Biomedicine,* vol. 141, pp. 19–26, Apr. 2017, doi: 10.1016/j.cmpb.2017.01.004

[32] D. Tomar and S. Agarwal, "Feature selection based least square win support vector machine for diagnosis of heart disease," *Int. J. Adv. Sci. Technol.,* vol. 65, pp. 39–58, 2014, doi: 10.14257/IJBSBT.2014.6.2.07

[33] H. Das, B. Naik, and H. S. Behera, "Medical disease analysis using neuro-fuzzy with feature extraction model for classification," *Informatics in Medicine Unlocked,* vol. 18, 2020, doi: 10.1016/j.imu.2019.100288

[34] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2923707

About the Authors

Tarun Kumar Agrawal is pursuing Ph.D. from Vivekananda Global University, Jaipur. He received M.Tech. (Computer Engineering) degree from Poornima University, India in 2016 and B.Tech. (CSE) degree from Rajasthan Technical University Kota, India. He is Assistant Professor with the Department of Computer Science at Vivekananda Global University, Jaipur. He has a total experience of 8 years in the field of teaching and research with some significant contributions in research. His current research interests are ML, AI, networking, Data Science, and Big Data.

Hemant Sharma received M.Tech. (Computer Engineering) degree from MNIT, India in 2014 and B.E. (Information Technology) degree from the University of Rajasthan, India. He is Assistant Professor with the Department of Computer Science at Vivekananda Global University, Jaipur, Rajasthan, India. He has a total experience of 12 years in the field of teaching and research with some significant contributions in research. His current research interests are Machine Learning, Artificial Intelligence, IoT, Semantic Web, Data Science, and Big Data.

Challa Madhavi Latha received Ph.D. from Vignan's Foundation for Science, Technology & Research, Guntur, Andhra Pradesh. She is working as Assistant Professor with the Department of CSE, CMR College of Engineering & Technology, Kandlakoya, Hyderabad, Telangana. She has more than 25 publications in various international journals and conferences. Her research areas include Capital Asset Pricing, Dynamic changes in the Stock market, and Stockholders' interests. In addition to Internet of Things, Cyber Security, Machine learning, and Cloud computing are the major areas of her research.

Sitaram Gupta received M.Tech (Computer Engineering) degree from RTU, Kota India in 2014 and B.E. (CSE) degree from MBM Jodhpur, India. He is HoD with the Department of Computer Science at Vivekananda Global University, Jaipur, Rajasthan, India. He has a total experience of 12 years in the field of teaching and research with some significant contributions in research. His current research interests are Machine Learning, Artificial Intelligence, IoT, Semantic Web, Data Science, and Big Data.