IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

ANALYSIS OF DIFFERENT ALGORITHMS FOR AUTOMATIC HEART DISEASE PREDICTION

[1]Shital V Pawar, [2] Dr. Bharat Gupta, [3]Dr. Jagdish D. Deshpande

[1]Research Scholar, [2]Professor, [3]Professor

[1]Department of Electronics.

[1]Shri JJT University, Rajasthan, India.

Abstract: Cardiovascular disease is a leading global killer. There are many individuals all around the globe who are afflicted with this fatal illness. The importance of early detection of heart disease becomes clear when looking at mortality rates and the enormous number of individuals who suffer from it. For this kind of sickness, the standard methods of diagnosis are insufficient. Clinical data analysis has several subfields, but one of the most prominent is cardiovascular disease prediction. The healthcare sector generates a massive quantity of data. Building a machine learning-based system for cardiac disease prediction improves diagnostic accuracy over current practices. The healthcare business generates vast amounts of data, and machine learning (ML) has been shown to be useful in generating judgments and predictions from this data. Predicting cardiac disease using ML approaches is only somewhat illuminated by the available research. In this study, we present a strategy for predicting cardiovascular illness with greater precision by using machine learning approaches to identify variables with high predictive value. Here, we present the prediction model by using a number of feature combinations and well-established categorization methods. By combining traditional statistical methods with a more modern hybrid approach, we are able to create a heart disease prediction model with improved performance and high accuracy.

***Index Terms*-heart disease prediction,long short-term memory, tasmanian devil optimization, deep learning, recurrent neural networkand performance metrics.**

I. INTRODUCTION

The terms "heart disease" and "cardiovascular disease" are often used synonymously with one another. "heart disease" In general, when people talk about cardiovascular disease, they are referring to illnesses that include constricted or blocked blood arteries, which may result in a heart attack, chest discomfort (also known as angina), or stroke. Heart-related diseases, also known as cardiovascular diseases (CVDs), have been the primary cause of a significant number of deaths around the world over the past few decades. Furthermore, heart-related diseases have emerged as the most life-threatening disease, not only in India, but all over the world. Therefore, there is a need for a system that is dependable, accurate, and practical in order to detect such disorders in a timely manner so that appropriate treatment may be administered. The algorithms and methods of machine learning have been applied to a variety of medical datasets in order to automate the study of extensive and complicated data sets.

In recent years, a great number of researchers have begun using a variety of machine learning strategies in an effort to support the medical community and professionals in their efforts to accurately diagnose heart-related conditions. Because of the various risk factors that may contribute to heart disease, such as diabetes, high blood pressure, excessive cholesterol, an irregular pulse rate, and many more, it can be difficult to diagnose heart disease in its early stages. In humans, the severity of cardiac disease has been determined using a number of different methods drawn from the fields of data mining and neural networks. Because of the complicated nature of cardiac disease, managing the condition effectively requires great attention. If you don't do it, it might hurt your heart or perhaps lead to an early death. Predictions of heart illness may be made using a variety of symptoms, including pulse rate, gender, age, and many more.

The primary goal of the study is to increase the performance accuracy of heart disease prediction by combining two models that offer greater accuracy. This will be accomplished by combining two models that give better accuracy. This study analyses the performance of a number of different models that are based on different machine learning and deep learning algorithms and approaches. It gives comparison findings of these models. The researchers have observed that models that are based on supervised learning algorithms such as SVM, Naive Bayes, Decision Trees, Random Forest, and ensemble models are particularly popular. In addition to this, we incorporated genetic algorithms and deep learning systems. The concept of a hybrid methodology, often known as the combining of several machine learning methods with a weighted average

The supplementary part of the article may be arranged ahead of time in the following manner: section 2 provided the earlier heart disease prognosis is supplied. The suggested model for the system is laid out and discussed in section 3. The part 4 contains a depiction of the results obtained using the approach that was planned. A synopsis of the piece may be found in section 5, which is provided for your convenience.

II. RELATED WORKS

The literature has given deep learning algorithms for heart disease prediction a lot of attention.

Youness et al, have made use of the Fast Correlation-Based Feature Selection (FCBF) technique to eliminate superfluous features and boost heart disease classification accuracy. We then use particle swarm optimization (PSO) and ant colony optimization (ACO) to classify data using a variety of classification techniques, including K-Nearest Neighbour, Support Vector Machine, Naive Bayes, Random Forest, and a Multilayer Perception | Artificial Neural Network. Applying the suggested hybrid technique to the heart disease dataset, the findings show the method's effectiveness and resilience in processing different kinds of data for illness classification. Therefore, this research evaluates the outcomes of several machine learning algorithms based on a variety of metrics (such as accuracy, precision, recall, f1-score, etc.). The initial stage of automated learning for cardiac illness detection, with potential for expansion into other areas of study. This research is limited in many ways: first, by the author's level of expertise; second, by the resources at hand (such as the speed of the computer utilized in this research); and third, by the time frame in which the research could be conducted.

Nimai Chand et al, data from 1094 people throughout India who were administered medication for a variety of conditions. We have developed a model that learns from this data and then attempts to predict the likelihood of a heart attack in unstamped data. This concept may facilitate shared decision making between the doctor and patient for optimal care. The model's True Positive Rate and False-Negative Rate, as well as its Area under the Receiver Operating Curve (AUC-ROC), must be taken into account in the validation set of data. When a patient has a high likelihood for a heart attack but the model predicts that there is no cardiac trouble, we want to reduce and punish the model for making a faulty forecast. As a result, researchers have proposed a new statistic they call Selection Value that accounts for such cases and chooses the method that yields the highest S.V. We attempt to provide the doctor with the best possible course of action based on the patient's medical history without influencing his or her decision in any way, as described in the hypothetical scenario section.

Amin et al,have created a heart disease prediction system based on machine learning and a dataset including information on cardiovascular illness. Metrics such as classification accuracy, specificity, sensitivity, Matthews' correlation coefficient, and execution time were used to assess the effectiveness of these seven well-known machine learning algorithms, three feature selection algorithms, the cross-validation method, and seven classifiers. A key feature of the proposed system is its ability to reliably distinguish between patients with and without cardiac disease. A receiver optimistic curve and area under the curve were also calculated for each classifier. Preventing and treating heart failure relies on early and accurate identification of heart disease. For many years, doctors have known that relying only on a patient's medical history to diagnose heart disease was flawed. Non-invasive techniques, such as machine learning, may accurately categorize individuals into those without cardiac disease and those with it. Making use of machine learning in the development of a decision support system for cardiac illness diagnostics is the optimal course of action. Furthermore, the performance of the diagnostic system was lowered and the calculation time was raised due to several useless aspects.

Senthil et al, have proposed a new strategy for predicting cardiovascular illness with greater precision via the use of machine learning algorithms to identify variables with high predictive value. Assorted feature combinations and well-established categorization methods are provided as building blocks of the prediction model. The healthcare business generates vast amounts of data, and machine learning (ML) has proven useful in generating judgments and predictions from this data. Predicting the occurrence of heart disease is a crucial yet difficult medical task. However, if the condition is diagnosed early and preventive measures are implemented promptly, the death rate may be dramatically reduced. It would be ideal to broaden the scope of this research to include real-world datasets rather than only theoretical methods and simulations. The suggested HRFLM method combines elements of both the Random Forest (RF) and the Linear Method (LM). When it came to predicting cardiovascular disease, HRFLM performed well.

Ganesan et al,have constructed a reliable methodology for predicting people who may acquire heart disease by combining data from the UCI Repository with information from healthcare sensors. In addition, patient data is classified using classification algorithms to help in the diagnosis of cardiac disease. The benchmark dataset will be used to train the classifier in the training phase. Real patient data is utilized during testing to confirm the diagnosis of illness. Several classifiers, including J48, logistic regression (LR), multilayer perception (MLP), and support vector machine (SVM), are put to the test on a benchmark dataset (SVM). All of the simulation results (accuracy, precision, recall, F-score, and kappa value) reveal that the J48 classifiers provide the best overall performance. A new Cloud and IoT based Healthcare application is designed to monitor and detect dangerous illnesses in order to provide excellent service to the user using the online healthcare services.

III. PROPOSED TECHNIQUE

The Cleveland UCI archive of heart disease cases has been classified using Python and is now part of the planned system. The dataset is represented visually, and a workspace and tools for developing predictive analytics are made available. The ML procedure starts with cleaning and organizing the input data, then moves on to DT entropy-based feature selection, an assessment of the model's performance, and finally the production of more precise results. For all possible permutations of characteristics, the process of picking features and modeling them will go on forever. Each model's iterative performance and the specific machine learning techniques used to achieve that performance are recorded. These models are constructed using a total of 13 features.

3.1 Input Dataset:

The data on heart disease is preprocessed once it has been compiled from a large number of records. The collection contains information on 303 patients, with 6 instances containing incomplete data. These 6 patient records were eliminated from the dataset, and the remaining 297 are being utilized in the project's pre-processing. Multiclass variables and binary classification are presented to categorize the features of the provided dataset. In order to ascertain whether or not a patient suffers from heart illness, the multi-class variable must be used. Patients who are experiencing heart problems have a value of 1, whereas those who are not experiencing heart problems have a value of 0. As part of the data-gathering process, medical records must be transformed into diagnostic values. Preprocessing data from 297 patient records yielded the following results: 137 records had a value of 1, indicating the existence of heart sickness, whereas the remaining 160 records had a value of 0, indicating the absence of heart disease.

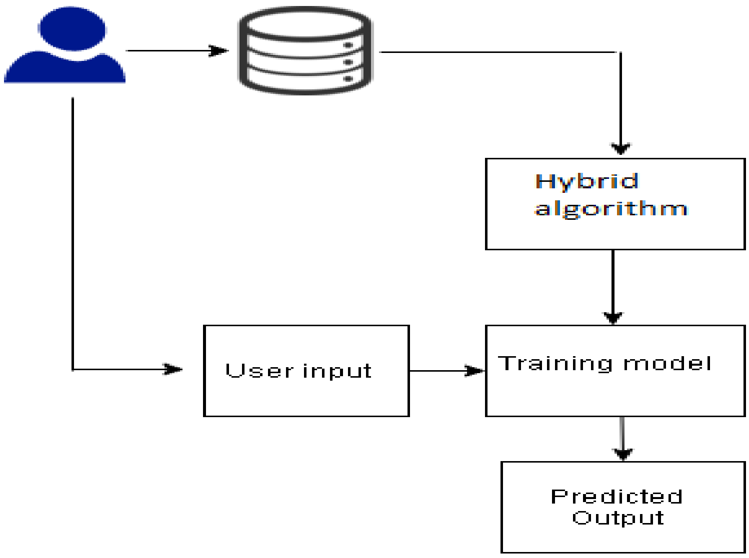

Fig 1: System Architecture

Predicting cardiovascular illness using machine learning and an automated medical diagnostic approach is the focus of this research. The hybrid model is used because it provides the most accurate results for classifying cardiovascular disease. The probabilities from one machine learning model are fed into another in a revolutionary approach known as a hybrid model.
When both machine learning methods are taken into account, as they are in our hybrid model, we get the best possible optimal outcomes. The suggested approach uses a novel-based hybrid model to anticipate cardiac issues using automated machine learning. Predictions of cardiovascular disease are made using this hybrid model. In this case, the Cleveland dataset is processed. Researchers in the field of machine learning often investigate this dataset. There are a total of 303 cases and around 14 features in this collection. The purpose of this research is to use a simple zero-to-one categorization, where "0" denotes the lack of heart illness and "1" indicates the presence of heart disease (present of heart disease). Based on the results given by our suggested model, patients may seek therapy. The suggested software is useful for anticipating patients' needs and acting accordingly.

IV. RESULTAND DISCUSSION

Accuracy, precision, and classification error are only few of the main performance measures that were used to determine this model's efficacy. In this context, accuracy would be the percentage of the training data that correctly predicts the target. Predictions that turn out to be true positives are one indicator of how precise a model is. The categorization error is the proportion of a given collection of instances that is either inaccurate or incomplete. Results from an HRFLM analysis. The ML technique allows you to test out several models and choose the one with the best results. For the prediction of cardiac problems, we provide HRFLM, a model with low classification error and excellent accuracy. Each classifier's results are examined independently, and the collected data is saved securely for further examination.
Every statistical metric is analyzed and the performance of the suggested strategy is calculated using the aforesaid methods. In this part, we exhibit graphical representations of the statistical measures made by both the proposed approach and the current method.

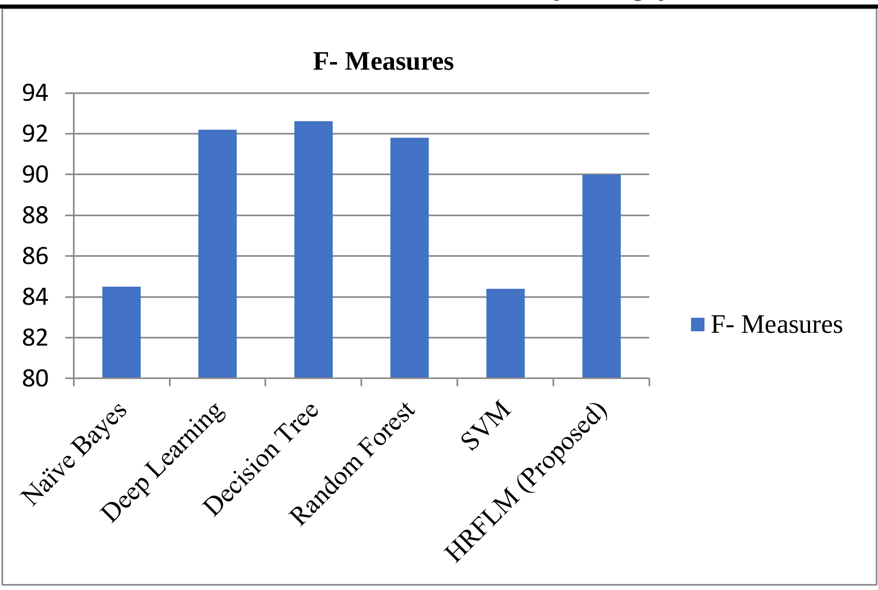

Fig 2: Performance evaluation of the accuracy

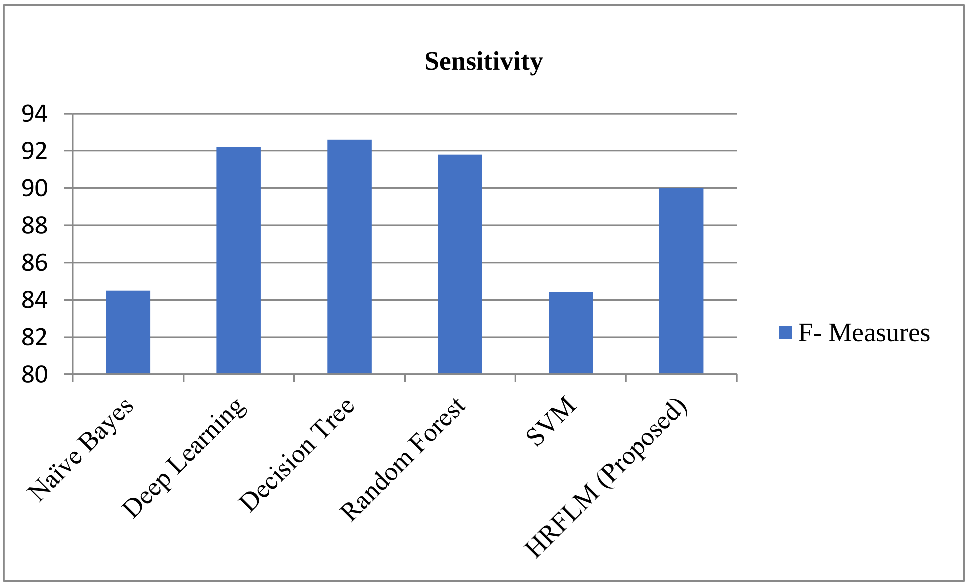

Fig 3: Performance evaluation of sensitivity

When compared to the suggested system, the other algorithms have lower accuracy. The suggested technique offers a superior answer to the problem of heart attack prediction over the currently available methods. Validating the effectiveness of the suggested strategy relies heavily on the projected accuracy metrics. Each technique is around 88.4% accurate. Among them, the suggested approach obtains an overall accuracy of just 88.4 percent. It has been shown via validation that the recommended approach yields the best outcomes.

Figure 3 depicts the next step in the statistical analysis of the sensitivity measurement. The objective of the suggested approach is a sensitivity of 92.8%. Also examined in that chart is the degree to which the current approach is sensitive. Naive Bayes, Deep Learning, Decision Tree, Random Forest, and Support Vector Machine all have lower sensitivity as a result. The suggested approach has achieved the best result in fault detection of defense sector in aviation and submarine systems compared to current approaches.

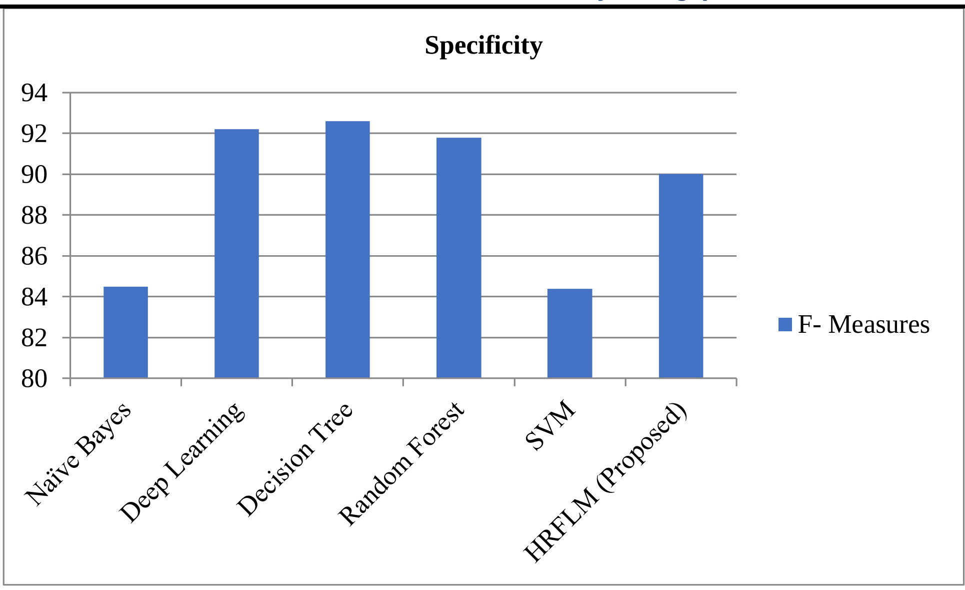

Fig 4: Performance evaluation of specificity

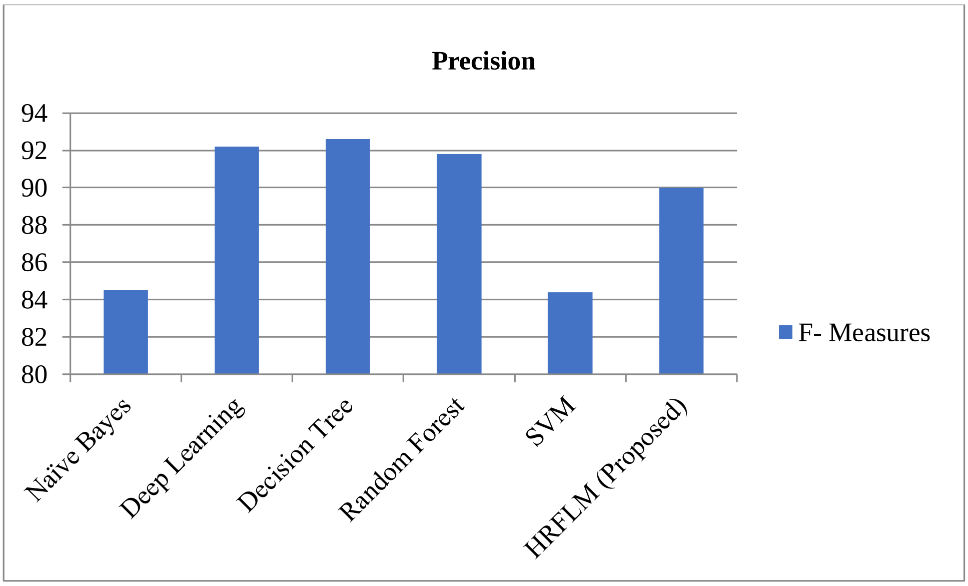

Fig 5: Performance evaluation of Precision

Figures 4 show measurements of specificity and accuracy made using the suggested technique and other methods. The suggested strategy has an impressive 82.6% specificity. Naive Bayes, Deep Learning, Decision Tree, Random Forest, and Support Vector Machine all have lower sensitivity than the suggested approaches.

Figure 5 shows that after evaluating the accuracy of the suggested technique, the confidence interval is 90.9%. The accuracy of Naive Bayes, Deep Learning, Decision Tree, Random Forest, and Support Vector Machine is lower than that of currently used approaches.

V. CONCLUSION

Techniques from the field of machine learning were used in this research project in order to analyze the raw data and offer a new and unique interpretation with regard to heart disease. The prediction of heart disease is a challenging endeavor that is well recognized for its importance within the medical profession. However, the mortality rate may be greatly lowered if the illness is recognized while it is still in its early stages and preventive measures are put into place as soon as it is practically possible to do so. It is strongly advised that this work be developed further in order to target the findings toward real-world datasets rather than only relying on theoretical methodologies and simulations. An HRFLM technique that combines the elements of Random Forest (RF) and Linear Method (LM) is devised and then put into practice (LM). It has been shown that the HRFLM is highly accurate in predicting the onset of heart disease. [Citation needed] This study might go in a number of different directions in the future, and each of those directions could be pursued using a different set of machine learning methods, ranging from improved prediction methods to a wide range of machine learning methods mixed together. Furthermore, it is conceivable to construct new feature selection algorithms in order to acquire a bigger perception of the relevant aspects and, as a result, increase the efficiency of heart disease prediction systems. This would be accomplished by expanding the scope of the features being considered.

REFERENCES

[1] S. Abdullah and R. R. Rajalaxmi, ``A data mining model for predicting the coronary heart disease using random forest classifier," in Proc. Int. Conf. Recent Trends Comput.Methods, Commun.Controls, Apr. 2012, pp. 22_25.

[2] H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar ``Using PSO algorithm for producing best rules in diagnosis of heart disease," in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306_311.

[3] N. Al-milli, ``Backpropogation neural network for prediction of heart disease," J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131_135, 2013.

[4] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, ``Analysis of neural networks based heart disease prediction system," in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233_239.

[5] 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, Jan. 2012. doi: 10.1016/j.jksuci.2011.09.002.

[6] L. Baccour, ``Amended fused TOPSIS-VIKOR for classi_cation (ATOVIC) applied to some UCI data sets," Expert Syst. Appl., vol. 99, pp. 115_125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025.

[7] H. A. Esfahani and M. Ghazanfari, ``Cardiovascular disease detection using a new ensemble classi_er," in Proc. IEEE 4th Int. Conf. Knowl.- Based Eng. Innov. (KBEI), Dec. 2017, pp. 1011_1014.

[8] F. Dammak, L. Baccour, and A. M. Alimi, ``The impact of criterion weights techniques in TOPSIS method of multi-criteria decision making in crisp and intuitionistic fuzzy domains," in Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE), vol. 9, Aug. 2015, pp. 1_

[9] S. M. S. Shah, S. Batool, I. Khan, M. U. Ashraf, S. H. Abbas, and S. A. Hussain, ``Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis," Phys. A, Stat. Mech. Appl., vol. 482, pp. 796_807, 2017. doi: 10.1016/j.physa.2017.04.113.

[10] Y. E. Shao, C.-D. Hou, and C.-C. Chiu, ``Hybrid intelligent modelling schemes for heart disease classification," Appl. Soft Comput. J., vol. 14, pp. 47_52, Jan. 2014. doi: 10.1016/j.asoc.2013.09.020.

[11] J. S. Sonawane and D. R. Patil, ``Prediction of heart disease using multilayer perceptron neural network," in Proc. Int. Conf. Inf. Commun. Embed- ded Syst., Feb. 2014, pp. 1_6.

[12] C. Sowmiya and P. Sumitra, ``Analytical study of heart disease diagnosis using classi_cation techniques," in Proc. IEEE Int. Conf. Intell. Techn. Control, Optim. Signal Process. (INCOS), Mar. 2017, pp. 1_5.

[13] B. Tarle and S. Jena, ``An arti_cial neural network based pattern classi_cationalgorithm for diagnosis of heart disease," in Proc. Int. Conf. Comput., Commun., Control Automat. (ICCUBEA), Aug. 2017, pp. 1_4.

[14] V. P. Tran and A. A. Al-Jumaily, ``Non-contact Doppler radar based prediction of nocturnal body orientations using deep neural network for chronic heart failure patients," in Proc. Int. Conf. Elect. Comput. Technol. Appl. (ICECTA), Nov. 2017, pp. 1_5.

[15] K. Uyar and A. Ilhan, ``Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks," ProcediaComput. Sci., vol. 120, pp. 588_593, 2017.

[16] T. Vivekanandan and N. C. S. N. Iyengar, ``Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease," Comput. Biol. Med., vol. 90, pp. 125_136, Nov. 2017.

[17] S. Radhimeenakshi, ``Classification and prediction of heart disease risk using data mining techniques of support vector machine and artificial neural network," in Proc. 3rd Int. Conf. Comput. Sustain. Global Develop. (INDIACom), New Delhi, India, Mar. 2016, pp. 3107_3111.

[18] R. Wagh and S. S. Paygude, ``CDSS for heart disease prediction using risk factors," Int. J. Innov. Res. Comput., vol. 4, no. 6, pp. 12082_12089, Jun. 2016.

[19] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, ``An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction," Expert Syst. Appl., vol. 68, pp. 163_172, Feb. 2017.

[20] S. Zaman and R. Tou_q, ``Codon based back propagation neural network approach to classify hypertension gene sequences," in Proc. Int. Conf. Elect., Comput. Commun. Eng. (ECCE), Feb. 2017, pp. 443_446.

[21] W. Zhang and J. Han, ``Towards heart sound classification without segmentation using convolutional neural network," in Proc. Comput. Cardiol. (CinC), vol. 44, Sep. 2017, pp. 1_4.

[22] Y. Meidan, M. Bohadana, A. Shabtai, J. D. Guarnizo, M. Ochoa, N. O. Tippenhauer, and Y. Elovici, ``Pro_lIoT: A machine learning approach for IoT device identification based on network traffic analysis," in Proc. Symp. Appl. Comput., Apr. 2017, pp. 506_509.

[23] J.Wu, S. Luo, S.Wang, and H.Wang, ``NLES: A novel lifetime extension scheme for safety-critical cyber-physical systems using SDN and NFV," IEEE Internet Things J., no. 6, no. 2, pp. 2463_2475, Apr. 2019.

[24] J. Wu, M. Dong, K. Ota, J. Li, and Z. Guan, ``Big data analysis-based secure cluster management for optimized control plane in software-defined networks, IEEE Trans. Netw. Service Manag., vol. 15, no. 1, pp. 27_38, Mar. 2018.

[25] J. Wu, M. Dong, K. Ota, J. Li, and Z. Guan, ``FCSS: Fog computing based content-aware filtering for security services in information centric social networks," IEEE Trans. Emerg. Topics Comput., to be published. doi: 10.1109/TETC.2017.2747158.

[26] G. Li, J. Wu, J. Li, K. Wang, and T. Ye, ``Service popularity-based smart resources partitioning for fog computing-enabled industrial Internet of things," IEEE Trans. Ind. Informat., vol. 14, no. 10, pp. 4702_4711, Oct. 2018.

[27] J. Wu, K. Ota, M. Dong, and C. Li, ``A hierarchical security framework for defending against sophisticated attacks on wireless sensor networks in smart cities," IEEE Access, vol. 4, pp. 416_424, 2016.