International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 20 August 2023 **Revised:** 08 October 2023 **Accepted:** 27 October 2023*

Heart Disease Detection by Machine Learning System

Dr. R. Sundar[1], Dr. K. Maithili[2], T. Raghavendra Gupta[3], P L Srinivasa Murthy[4], G. Nagarjuna Rao[5]
[1]Assistant Professor,Computer Science and Engineering,Madanapalle Institute of Technology & Science Post Box No: 14, Kadiri Road Angallu (V), Madanapalle-517325Annamayya District, Andhra Pradesh, India.
Mail :drsundarr@mits.ac.in
[2]Associate Professor Department of CSE KG REDDY COLLEGE OF ENGINEERING &TECHNOLOGY
Moinabad, Hyderabad, Telangana-501504 Mail : drmaithili@kgr.ac.in
[3]Associate Professor Department of computer science and Engineering Hyderabad Institute of Technologgy And Management Hyderabad, Telangana Mail : raghu.ht@gmail.com
[4]Professor, Department of Computer science and Engineering, IARE Mail : plsrinivasamurthy@iare.ac.in
[5]Assistant Professor Department of CSE MLR Institute of Technology, Dundigal, Hyderabad
Mail : nagarjunarao.gudelli@gmail.com

Abstract: Heart disease is a prevalent global health issue that impacts a substantial number of individuals worldwide. It is characterized by symptoms such as shortness of breath, muscle weakness, and swollen feet. However, the current diagnostic methods for heart disease have limitations in terms of accuracy and efficiency, making early detection challenging. Consequently, researchers are striving to develop an effective approach for early detection of heart disease. The lack of advanced medical equipment and qualified healthcare professionals further complicates the diagnosis and management of cardiac conditions., there have been approximately 26 million reported cases of heart disease, with an additional 3.6 million new cases identified annually. In the United States, a significant proportion of the population is affected by heart disease. Typically, doctors diagnose heart disease by considering the patient's medical history, conducting a physical examination, and assessing any concerning symptoms. However, this diagnostic method does not consistently provide accurate identification of individuals with heart disease. The importance of employing. There are numerous crucial elements in the process for developing a smart parking system in an IoT context. First, sensors are placed in parking places to gather up-to-the-minute occupancy information. Then, using wireless communication protocols, this data is sent to a central server or cloud computing platform. After that, a data processing and analysis module interprets the gathered data using algorithms and machine learning techniques and presents parking availability information to users via a mobile application or other user interfaces. For effective management and monitoring of parking spaces, the system also includes automated payment methods and interacts with existing infrastructure. “Patient 1,patient 2,patient 3 and patient 4.” Dyspnea can be described as a sensation of breathlessness and inadequate breathing, where one feels unable to take in enough air or breathe deeply. It involves the interplay of mechanoreceptors in the upper airways, lungs, and chest wall, along with peripheral receptors, chemoreceptors, and other sensory receptors. Edema refers to the accumulation of excessive fluid in the body tissues, leading to swelling. While edema can occur in any part of the body, it is more commonly observed in the lower extremities Ascites - The pathological buildup of fluid in the abdominal cavity is known as ascites. It is the most frequent cirrhosis consequence and happens in 50% of patients with decompensated cirrhosis within 10 years. Ascites formation marks the change from stressed to decompensated cirrhosis. Patent 1 is in rank 1 and patient 5 is ranked 5. In weighted table every value is equally split by 1,so that each value is equal. In the study, the researchers compared the sensitivity levels of two classifiers: the Relief FS method with a linear SVM classifier and the NB classifier with specific features from the LASSO FS algorithm. The findings revealed that the NB classifier, utilizing LASSO FS features, exhibited the highest performance in terms of sensitivity. Additionally, the Logistic Regression MCC classifier, employing the FCMIM FS method, achieved a classification accuracy of 91%.

Keywords: Heart Disease, algorithms, artificial..

1. INTRODUCTION

Addressing the challenges of cost and computational complexity associated with analysis [8], there is a need to develop a noninvasive diagnostic approach utilizing machine learning (ML) algorithms. In this context, researchers [11] and [12] have employed the Cleveland cardiovascular disease dataset to tackle the problem of identifying heart disease (HD). To effectively train and test ML prediction models, it is important to have appropriate data. Balanced datasets can improve .In addition; an important aspect of improving model predictions is the incorporation of enhanced features derived. Therefore, the process of becomes crucial in enhancing the predictive capabilities of the models. Although several diagnostic methods have been proposed in the literature, they do not consistently provide reliable diagnosis of high definition (HD). As a result, it is necessary to carefully evaluate and choose appropriate techniques for data preprocessing and feature engineering. To optimize the accuracy and performance of a model, it is essential to preprocess the data effectively and select relevant features. This helps enhance the quality of the input data and improves the model's ability to make accurate predictions or outcomes. Data preprocessing involves utilizing various techniques to prepare the data for machine learning models Some commonly used techniques for data preprocessing and feature selection include Standardized Scalar (SS) for scaling features, Min-Max Scalar for normalizing features, and handling missing values by eliminating them from the dataset. Feature selection plays a crucial role in improving model performance, and various methods are available for this purpose. These methods include: Least Absolute Shrinkage Selection Operator) Relief Minimal Redundancy Maximum Relevance Local Learning Based Features Selection Principal Component Analysis Optimization methods like Ant Colony Optimization), Fruit Fly Optimization and Bacterial Foraging Optimization These techniques help select the most relevant and informative features for the model. They cater to different types of feature selection tasks, such as secure feature selection, selecting features for high-dimensional small sample data, and handling large-scale data. By using these techniques effectively, we can improve prediction accuracy and enhance overall model performance. As the field of feature selection has advanced, researchers have explored various topics, including adversarial feature selection, internet feature selection, dispersed feature selection, multi-view feature selection, and stable feature selection. Jindong et al. have discussed the challenges associated with feature selection for large datasets. To summarize, data preprocessing techniques like data standardization and handling missing values, as well as feature selection methods, are essential for improving the performance of machine learning models across various data types and addressing specific challenges in feature selection. The curse of dimensionality necessitates reducing the complexity of data for different learning tasks. The selection of features greatly impacts applications such as simplifying buildings, enhancing learning outcomes, and generating clear and understandable data. However, due to the high dimensionality of big data, feature selection becomes a challenging task that introduces significant difficulties arise when performing feature selection on structured, heterogeneous, and streaming data, posing challenges related to stability and scalability. Resolving these feature selection issues is crucial in the context of big data analytics. To address these constraints, researchers developed an unsupervised hashing algorithm called topic hyper network hashing, as described in [15]. This algorithm leverages the contextual information from supplemental words surrounding images, effectively overcoming the semantic limitations of other hashing algorithms. Topic hyper network hashing is particularly well-suited for mobile image retrieval and has the potential to outperform several state-of-the-art techniques. . The filter approach is computationally simpler compared to the wrapper method. Additionally, the features selected through the filter approach are universally applicable to all models and are not specific to any particular model In this research, the focus was on the significance of global relevance when selecting features. The objective was to develop a diagnostic technique using machine learning to identify Huntington's disease Various machine learning prediction models such as artificial neural network logistic regression k-nearest neighbors), support vector machine decision tree and naive Bayes were employed for this purpose. To determine the relevant features, advanced algorithms like, Local-learning-based feature selection), and the fast conditional mutual data approach were utilized. The selection of the best model was done by optimizing hyper parameters through the Leave-One-Subject-Out Cross-Validation (method.. The Cleveland HD dataset was employed to test the proposed approach. However, it is essential to note that effective machine learning requires an appropriate model. A strong machine learning model not only performs well on the training data but also on unseen data. Hence, it is crucial to assess the classifiers using different data and determine their average correct classification rate. In this particular case, the average correct classification rate was found to be 50% [16].

2. MATERIALS AND METHODOLOGY

Dyspnea can be defined as the condition characterized by breathlessness or difficulty in breathing. It is a symptom commonly associated with a range of medical conditions, including respiratory and cardiac disorders. Edema refers to the swelling that occurs when there is an excess accumulation of fluid in the body's tissues. It can manifest in various areas of the body, such as the legs, ankles, and hands, and is often caused by underlying medical conditions like heart failure, kidney disease, or liver cirrhosis. Fatigue is a state of profound tiredness or exhaustion that significantly impacts a person's physical and mental abilities. It can be caused by a variety of factors, including insufficient sleep, chronic illnesses, stress, or certain medications. Ascites is the abnormal buildup of fluid within the abdominal cavity. It is commonly observed in conditions such as liver cirrhosis, liver failure, or certain types of cancer. Ascites can result in abdominal swelling and discomfort.The importance of applying machine learning algorithms for heart disease identification is extensive & has an opportunity to completely transform the healthcare industry. Here are a few important research implications. greater precision Huge to find complicated patterns that may not be visible to human specialists. These algorithms can be used by researchers to create models that could increase the precision of coronary artery disease identification and diagnosis. This may result in forecasts that are more accurate and trustworthy, with fewer false positives and negatives. Early identification or prevention: oblique risk factors that may escape the attention of medical experts. Early detection of these patterns allows for the implementation of therapies to stop the course of cardiac disease or lessen its effects, possibly preventing lives and saving money on healthcare. In the field of personalized medicine, machine learning algorithms are able to combine and examine huge datasets that include information on genetic, medical, lifestyle, and environmental aspects. Due

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *20 August 2023* ***Revised:*** *08 October 2023* ***Accepted:*** *27 October 2023*

to the abundance of data available, it is now possible to create personalized models that can determine a person's risk of developing heart disease and adjust their therapy appropriately. Such personalized medicine methods have the possibility to enhance patient outcomes and enhance the distribution of healthcare resources. Effective patient triage & resource allocation: By forecasting the seriousness and urgency of a patient's heart disease state, machine learning algorithms can help healthcare practitioners triage patients. This can aid in more efficient resource allocation and guarantee that patients receive the right care based on their unique risk profiles. Machine learning may lead to better care for patients and increased healthcare system effectiveness by optimizing resource allocation. Insights into the pathophysiology of diseases: Artificial intelligence algorithms can analyses intricate interconnections within massive datasets, revealing previously unrecognized links between different risk factors with heart disease. These discoveries may aid in the discovery of new biomarkers, new treatment targets, and improved understanding of the fundamental processes behind the development of cardiac disease. The velocity of studies and discoveries may be substantially accelerated by machine learning's capacity to take on large-scale data processing and develop ideas. Clinical decision support: Machine learning algorithms can help physicians make decisions by giving them recommendations and extra information based on the study of patient data. This can help physicians provide patients with better care and prevent medical mistakes by helping them make more informed and precise assessments, treatment plans, and treatments. In conclusion, machine learning systems' potential to improve accuracy, enable prevention and early detection, personalize medicine, optimize the use of resources, provide insights into the causes of disease, and support medical decisions make them important for the detection of heart disease. These developments hold the promise of greatly enhancing patient treatment, lowering healthcare expenses, and expanding our knowledge of cardiac disease.The WPM method, commonly referred to as "Words Per Minute," is a measure of typing speed. It is used to determine how many words a person can type in one minute. WPM is often used as a standard unit of measurement for typing speed in various contexts, such as data entry jobs, transcription work, and typing competitions.To calculate your typing speed in WPM, follow these steps:Select a passage or text sample that you will use for the typing test. It can be a paragraph from a book, an article, or any piece of text.Set a timer for one minute.Begin typing the selected passage as accurately and quickly as possible, without making significant errors or backspacing excessively.Stop typing when the timer goes off, even if you haven't finished the entire passage.Count the total number of words you have typed. You can use the word count feature in word processors or online tools to assist you.Your WPM score is calculated by dividing the total number of words typed by the elapsed time in minutes. For example, if you typed 80 words in one minute, your WPM score would be 80.It's worth noting that some typing tests may deduct words for errors or subtract time spent on correcting mistakes. This approach ensures accuracy and penalizes excessive errors. However, in simpler tests, errors might not be taken into account.To improve your typing speed, you can practice regularly by typing different texts or using online typing tutorials and games. Focusing on proper finger placement, using all fingers efficiently, and gradually increasing your speed can help you improve your WPM over time.

3. RESULT AND DISCUSSION

TABLE 1. Heart Disease Detection by Machine Learning System

Symptoms	Dyspnea	Edema	Fatigue	Ascites
Patient 1	0.87	0.72	0.55	0.42
Patient 2	0.37	0.72	0.87	0.55
Patient 3	0.22	0.37	0.37	0.59
Patient 4	0.72	0.37	0.87	0.42
Patient 5	0.42	0.22	0.72	0.59

Table 1 shows the table represent the severity or intensity of the symptom for each patient, with higher values indicating more severe symptoms.For example, Patient 1 has a dyspnea score of 0.87, an edema score of 0.72, a fatigue score of 0.55, and an ascites score of 0.42. Patient 2 has a dyspnea score of 0.37, an edema score of 0.72, a fatigue score of 0.87, and an ascites score of 0.55.This type of table can be used for various purposes, such as tracking symptom progression, comparing symptom severity among patients, or analyzing patterns and relationships between symptoms.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 20 August 2023 **Revised:** 08 October 2023 **Accepted:** 27 October 2023*

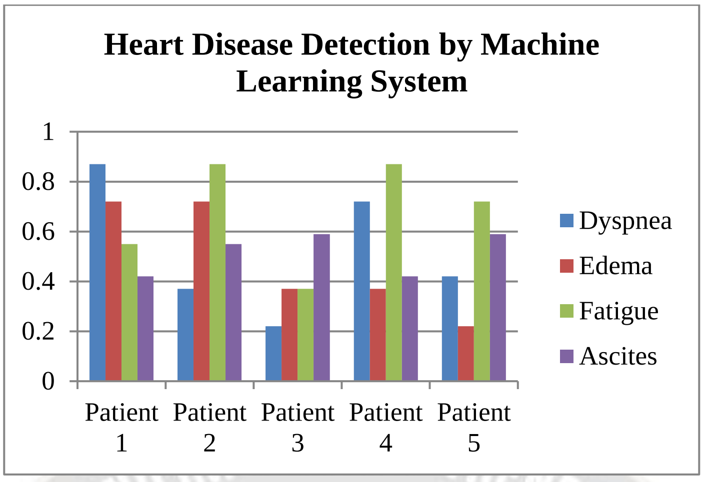

FIGURE 1. Heart Disease Detection by Machine Learning System data set

Figure 1 shows that thePatient 1 has a dyspnea score of 0.87, an edema score of 0.72, a fatigue score of 0.55, and an ascites score of 0.42. Patient 2 has a dyspnea score of 0.37, an edema score of 0.72, a fatigue score of 0.87, and an ascites score of 0.55.

TABLE 2. Performance value

Symptoms	Dyspnea	Edema	Fatigue	Ascites
Patient 1	1.00000	1.00000	0.67273	1.00000
Patient 2	0.42529	1.00000	0.42529	0.76364
Patient 3	0.25287	0.51389	1.00000	0.71186
Patient 4	0.82759	0.51389	0.42529	1.00000
Patient 5	0.48276	0.30556	0.51389	0.71186

Table 2 performance value Similar to the previous table, each row represents a different patient, and each column represents a different symptom. The numerical values in the table represent the severity or intensity of the symptom for each patient, with higher values indicating more severe symptoms.For example, Patient 1 has a dyspnea score of 1.00000, an edema score of 1.00000, a fatigue score of 0.67273, and an ascites score of 1.00000. Patient 2 has a dyspnea score of 0.42529, an edema score of 1.00000, a fatigue score of 0.42529, and an ascites score of 0.76364.This type of table can be used for symptom assessment, monitoring disease progression, or conducting statistical analyses related to symptom severity among patients.

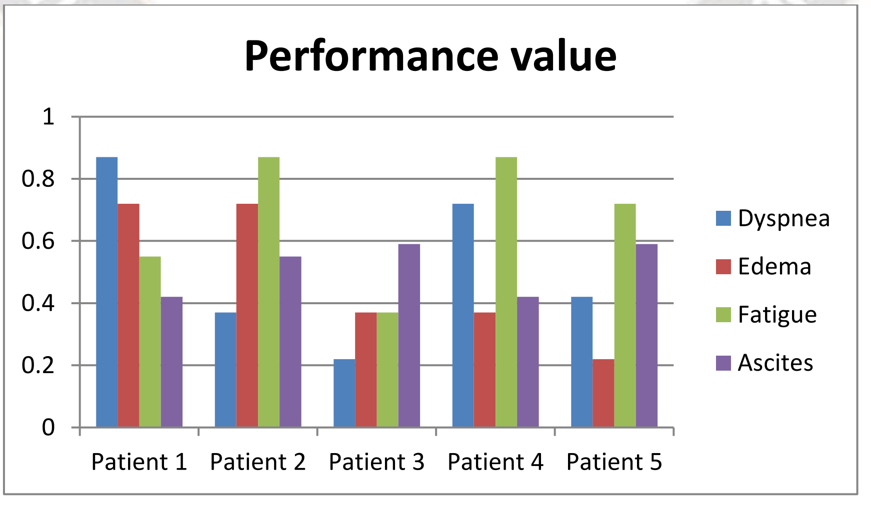

FIGURE 2. Performance value

Figure 2 showsPatient 1 has a dyspnea score of 1.00000, an edema score of 1.00000, a fatigue score of 0.67273, and an ascites score of 1.00000. Patient 2 has a dyspnea score of 0.42529, an edema score of 1.00000, a fatigue score of 0.42529, and an ascites score of 0.76364.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 20 August 2023 **Revised:** 08 October 2023 **Accepted:** 27 October 2023*

TABLE 3. Weightage

Symptoms	Dyspnea	Edema	Fatigue	Ascites
Patient 1	0.25	0.25	0.25	0.25
Patient 2	0.25	0.25	0.25	0.25
Patient 3	0.25	0.25	0.25	0.25
Patient 4	0.25	0.25	0.25	0.25
Patient 5	0.25	0.25	0.25	0.25

Table 3 shows that the Weightage are same values.

TABLE 4. Weighted normalized decision matrix

Weighted normalized decision matrix				
Patient 1	1.00000	1.00000	0.90565	1.00000
Patient 2	0.80755	1.00000	0.80755	0.93481
Patient 3	0.70913	0.84668	1.00000	0.91854
Patient 4	0.95379	0.84668	0.80755	1.00000
Patient 5	0.83355	0.74349	0.84668	0.91854

Table 4show the weighted normalized decision matrix for the given patients, we can identify the highest and lowest values in each row.Patient 1:Highest value: 1.00000Lowest value: 0.90565Patient 2:Highest value: 1.00000Lowest value: 0.80755Patient 3:Highest value: 1.00000Lowest value: 0.70913Patient 4:Highest value: 1.00000Lowest value: 0.80755 Patient 5:Highest value: 0.91854Lowest value: 0.74349

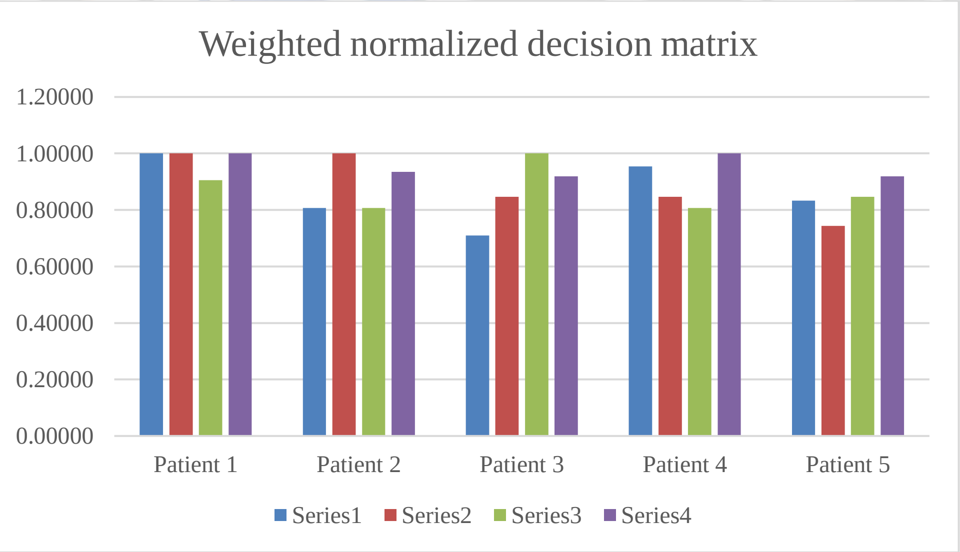

FIGURE 3. Weighted Normalized Decision Matrix

Figure 3 show the weighted normalized decision matrix for the given patients, we can identify the highest and lowest values in each row.Patient 1:Highest value: 1.00000Lowest value: 0.90565Patient 2:Highest value: 1.00000Lowest value: 0.80755Patient 3:Highest value: 1.00000Lowest value: 0.70913Patient 4:Highest value: 1.00000Lowest value: 0.80755 Patient 5:Highest value: 0.91854Lowest value: 0.74349.

TABLE 5. Final Result of Heart Disease Detection by Machine Learning System

	Preference Score	Rank
Patient 1	0.90565	1
Patient 2	0.60963	3
Patient 3	0.55150	4
Patient 4	0.65214	2
Patient 5	0.48197	5

Table 5 show that these ranks indicate the relative positions of the patients based on the numbers provided. Patient 1 has the lowest number and therefore holds the highest rank, while Patient 5 has the highest number and holds the lowest rank.

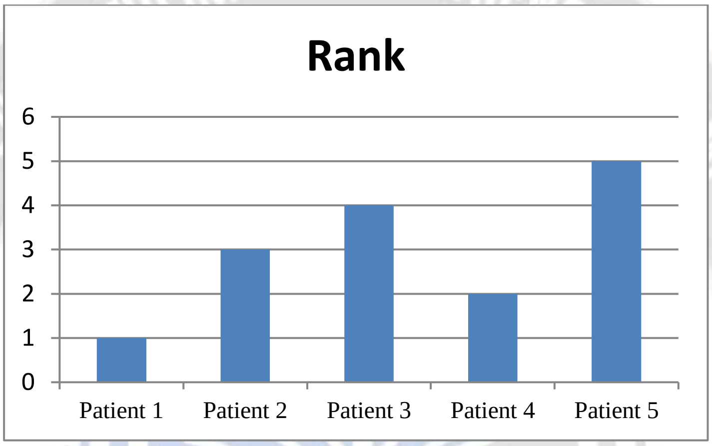

Figure 4. Ranking

Figure 4 shows that these ranks indicate the relative positions of the patients based on the numbers provided. Patient 1 has the lowest number and therefore holds the highest rank, while Patient 5 has the highest number and holds the lowest rank.

4. CONCLUSION

To address the challenge of selecting relevant features, the study employs four commonly used algorithms (Relief, MRMR, LASSO, LLBFS) and introduces a new method called FCMIM. The system determines the optimal hyperparameters through the LOSO cross-validation technique. Based on the findings presented in Table 15, the evaluation results indicate that the Artificial Neural Network (ANN) classifier.This suggests that the ANN classifier with Relief FS is the most effective technique for accurately identifying individuals who are healthy. On the other hand, the Naive Bayes (NB) classifier, when utilizing specific features obtained from the LASSO FS algorithm, demonstrates the highest sensitivity compared to the Relief FS method with the Support Vector Machine (SVM) classifier (linear kernel). In other words, the NB classifier with LASSO FS is particularly adept at correctly identifying positive cases. Lastly, the Logistic Regression MCC classifier, when employing the FCMIM FS method, achieves an impressive classification accuracy rate of 91%.

REFERENCES

[1]. A. L. Bui, T. B. Horwich, and G. C. Fonarow, ‘‘Epidemiology and risk profile of heart failure,’’ Nature Rev. Cardiol., vol. 8, no. 1, p. 30, 2011.
[2]. M. Durairaj and N. Ramasamy, ‘‘A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate,’’ Int. J. Control Theory Appl., vol. 9, no. 27, pp. 255–260, 2016.
[3]. L. A. Allen, L. W. Stevenson, K. L. Grady, N. E. Goldstein, D. D. Matlock, R. M. Arnold, N. R. Cook, G. M. Felker, G. S. Francis, P. J. Hauptman, E. P. Havranek, H. M. Krumholz, D. Mancini, B. Riegel, and J. A. Spertus, ‘‘Decision making in advanced heart failure: A scientific statement from the American heart association,’’ Circulation, vol. 125, no. 15, pp. 1928–1952, 2012.

[4]. S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, ''Innovative artificial neural networks-based decision support system for heart diseases diagnosis,'' J. Intell. Learn. Syst. Appl., vol. 5, no. 3, 2013, Art. no. 35396.

[5]. Q. K. Al-Shayea, ''Artificial neural networks in medical diagnosis,'' Int. J. Comput. Sci. Issues, vol. 8, no. 2, pp. 150–154, 2011.

[6]. J. Lopez-Sendon, ''The heart failure epidemic,'' Medicographia, vol. 33, no. 4, pp. 363–369, 2011.

[7]. P. A. Heidenreich, J. G. Trogdon, O. A. Khavjou, J. Butler, K. Dracup, M. D. Ezekowitz, E. A. Finkelstein, Y. Hong, S. C. Johnston, A. Khera, D. M. Lloyd-Jones, S. A. Nelson, G. Nichol, D. Orenstein, P. W. F. Wilson, and Y. J. Woo, ''Forecasting the future of cardiovascular disease in the united states: A policy statement from the American heart association,'' Circulation, vol. 123, no. 8, pp. 933–944, 2011.

[8]. A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, ''Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity,'' J. Roy. Soc. Interface, vol. 8, no. 59, pp. 842–855, 2011.

[9]. S. I. Ansarullah and P. Kumar, ''A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method,'' Int. J. Recent Technol. Eng., vol. 7, no. 6S, pp. 1009–1015, 2019.

[10]. S. Nazir, S. Shahzad, S. Mahfooz, and M. Nazir, ''Fuzzy logic based decision support system for component security evaluation,'' Int. Arab J. Inf. Technol., vol. 15, no. 2, pp. 224–231, 2018.

[11]. R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, ''International application of a new probability algorithm for the diagnosis of coronary artery disease,'' Amer. J. Cardiol., vol. 64, no. 5, pp. 304–310, Aug. 1989.

[12]. J. H. Gennari, P. Langley, and D. Fisher, ''Models of incremental concept formation,'' Artif. Intell., vol. 40, nos. 1–3, pp. 11–61, Sep. 1989.

[13]. Y. Li, T. Li, and H. Liu, ''Recent advances in feature selection and its applications,'' Knowl. Inf. Syst., vol. 53, no. 3, pp. 551–577, Dec. 2017.

[14]. J. Li and H. Liu, ''Challenges of feature selection for big data analytics,'' IEEE Intell. Syst., vol. 32, no. 2, pp. 9–15, Mar. 2017.

[15]. L. Zhu, J. Shen, L. Xie, and Z. Cheng, ''Unsupervised topic hypergraph hashing for efficient mobile image retrieval,'' IEEE Trans. Cybern., vol. 47, no. 11, pp. 3941–3954, Nov. 2017.

[16]. S. Raschka, ''Model evaluation, model selection, and algorithm selection in machine learning,'' 2018, arXiv:1811.12808. [Online]. Available: http://arxiv.org/abs/1811.12808

[17]. S. Palaniappan and R. Awang, ''Intelligent heart disease prediction system using data mining techniques,'' in Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl., Mar. 2008, pp. 108–115. [18] 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, p. 72, 2015.

[18]. 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, May 2009.

[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.