TANFIS Classifier Integrated Efficacious Aassistance System for Heart Disease Prediction using CNN-MDRP

O. Bhaskaru † and M. Sreedevi ††,

† Research Scholar, CSE Department, Koneru Lakshmaiah Education Foundation, Green Fields, Guntur District, Vaddeswaram, Andhra Pradesh, 522502.
†† Professor, CSE Department, Koneru Lakshmaiah Education Foundation, Green Fields, Guntur District, Vaddeswaram, Andhra Pradesh,522502.

Abstract

A dramatic rise in the number of people dying from heart disease has prompted efforts to find a way to identify it sooner using efficient approaches. A variety of variables contribute to the condition and even hereditary factors. The current estimate approaches use an automated diagnostic system that fails to attain a high level of accuracy because it includes irrelevant dataset information. This paper presents an effective neural network with convolutional layers for classifying clinical data that is highly class-imbalanced. Traditional approaches rely on massive amounts of data rather than precise predictions. Data must be picked carefully in order to achieve an earlier prediction process. It's a setback for analysis if the data obtained is just partially complete. However, feature extraction is a major challenge in classification and prediction since increased data increases the training time of traditional machine learning classifiers. The work integrates the CNN-MDRP classifier (convolutional neural network (CNN)-based efficient multimodal disease risk prediction with TANFIS (tuned adaptive neuro-fuzzy inference system) for earlier accurate prediction. Perform data cleaning by transforming partial data to informative data from the dataset in this project. The recommended TANFIS tuning parameters are then improved using a Laplace Gaussian mutation-based grasshopper and moth flame optimization approach (LGM^{2}G). The proposed approach yields a prediction accuracy of 98.40 percent when compared to current algorithms.

Keywords:
Convolution neural network (CNN), TANFIS, CNN-MDRP, Heart disease prediction, Accuracy, Clinical data analysis, data mining.

1. Introduction

Heart disease is the utmost common of decease in India and throughout. Conferring to the World Health Organization (WHO), cardiac disorders kill 17.7 million individuals globally every year, amounting for 31% of all deaths [1]. Predicting cardiac illness is a difficult undertaking, and doctors must devote a significant amount of effort to it. As a result, this is a critical moment to reduce the fatality rate by appropriately diagnosing the condition early on. The current prediction approaches use an automated diagnostic system that fails to attain a high level of accuracy because it includes irrelevant dataset information. To extract knowledge from datasets, data mining technologies can be utilised. As a result, it's critical to keep track of the most crucial symptoms that lead to cardiovascular disease (CVD).

The heart is perhaps the most vital organ. It simply controls blood flow across our bodies. Any wrongdoing in the heart can induce pain all over the body. An unhealthy lifestyle, such as smoking, excessive alcohol consumption, and excessive fat consumption, can all lead to hypertension [2]. Only a healthy lifestyle and early identification can avoid heart-related illnesses. The most pressing issue in modern healthcare is the provision of high-quality services and accurate analysis [3]. It really is one of the illnesses that can be addressed and regulated well. The accuracy of disease management is determined by the precise timing of sickness detection. The recommended investigation purposes to identify convinced cardiac conditions early to avoid significant repercussions. Therapeutic experts have accumulated a large amount of health data that may be evaluated and helpful information retrieved. Techniques for extracting meaningful and hidden information from massive volumes of data are known as data mining techniques. As a result, making judgments based on discrete data becomes challenging. Information retrieval sector Machine Learning (ML) excels at managing large datasets that have been well-formatted. The fundamental box of this research is to give doctors a means for detecting heart illness early on [4]. As a result, it will be simpler to adequately treat patients and avoid disastrous consequences. Machine learning's role in discovering hidden discrete patterns and analysing data is crucial. Machine learning techniques support in the forecast and early judgement of heart illness after data analysis.

Machine learning is getting increasingly popular as the quantity of data grows. Humans can learn from large volumes of data using machine learning, which is difficult and sometimes impossible for humans to accomplish [5]. It makes every effort to discover problems sooner, but there are still several traps that lead to incorrect diagnoses. The automated prediction is based on structured data, which is

Manuscript received October 5, 2022
Manuscript revised October 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.10.22

information and has been well utilizing data from such a dataset. The data utilised to anticipate cardiac illness includes ECG output, laboratory information, and health records from medical providers. The TANFIS classifier has been suggested in this study, which uses CNN-MDRP hybridization to extract features and classify heart illness. This work focuses on the data that is most important for prediction, removing unnecessary information using CNN-MDRP and improving the accuracy of selected features using LGM^2G algorithm. The TANFIS classifier was used to detect and categorize cardiac disease using the provided solution. To examine the utility of the suggested TANFIS-LGM^2G approach, different performance metrics were utilised.

The following are the main contributions of this article:

1. The CNN-MDRP method cleans the data by turning partial data into informative data out from dataset.
2. To predict heart illness, a innovative and effective TANFIS classifier is used. This prototype is reliable and proves to be a wonderful model for producing superior results than computational and mathematical methodologies.
3. The fundamental technique for optimizing the TANFIS parameters was the blended LGM^2G optimization.

The remainder of this article follows this format: The associated work is summarized in Section 2. The contribution is briefly discussed in Section 3. Section 4 backs up the findings and discussion. Section 5 brings the article to a close.

2. Related Works

The healthcare industry generates enormous amounts of data, which is unfortunately not "mined" for concealed information that would help with better decision-making. Often, hidden patterns and connections go unnoticed. This problem can be solved using advanced mining techniques.

To circumvent the limitations of machine learning, Vinitha et al. [6] developed the notion of machine learning-based illness prediction utilizing large data. This concept (a) eliminates missing information and (b) delivers accurate disease forecast. This method solves the present system's major two problems: both partial and missing data. The model of latent factors must be revised. The plan is to get the data from such a hospital that acquired it through a "structured and unstructured data" forum and analyse it using Machine Learning methods. For data partitioning, the MR method is used. With the typical pace, it reaches 94.8 percent, but it is faster than CNN-UDRP when it comes to reporting illness incidences.

The info mining idea "Disease Prediction by Machine Learning" was proposed by Sayali and Rashmi [7]. Any profession requires data analysis, and the event's optimum growth is incorporating that technique into the clinical foundation. The rapid rise of the medical care area is referred to as data mining, which forecasts information for healthcare. The current one is intended to explain the total healthcare systems by I analysing, (ii) managing, and (iii) forecasting healthcare data. In these types of procedures, machine learning is employed to retrieve illnesses information, and collected information is used to achieve therapeutic processes. The decision tree is used to anticipate disease outbreaks since it is extremely successful. The conclusion of this concept-based experiment is connected to disease symptoms, prompting the employment of a modified forecasting model to describe the data. If the approach selects a training set such as healthcare patient sensations, it will then employ the decision tree, forecast, and eventually provide the patient's symptoms in order to come up with a reliable sickness forecast. This approach is solely implemented, that is, it predicts only patient-related data in a short amount of time and at a cheap cost.

Diagnosed on the basis of clinical data, Otoom et al. [8] suggested an intelligent classifier forecast a heart disease issue. Cleveland Heart data was given by UCI. The majority of currently available mobile healthcare systems concentrate on data collection and surveillance, with little emphasis dedicated to real-time diagnosis. This diagnosis component is coupled with a real-time screening in the mobile application, which continually observes the patient and triggers an alarm if an emergency occurs. The cross-validation test revealed that the suggested diagnosis component was effective, with a performance metric accuracy of more than 85 percent.

Data mining algorithms for identifying cardiac disease were proposed by Chaurasia et al. [9]. The WEKA data mining tool is utilised, which includes a number of data mining approaches. 85.3 percent of the accuracy may be attributed to bagging. This data set's classification rate is improved by bagging.

Nagaraj M Lutimath and colleagues used Naive Bayes classification and SVM to predict cardiac disease. SVM outperforms Naive Bayes in terms of accuracy, according to several performance measures [10].

Vidhya et al. [11] projected a fresh modified ANFIS (M-ANFIS) to identify a variety of disorders. The healthcare data is handled in the preparation step, and the dataset's characteristics are extracted. The grouping may then be described using k-medoid segmentation, which determines the closed frequency item set. This grouping approach provides unequal grouping during extracting the features.

The basic principle behind the recommended method was to create a medical decision support system based necessary input after reviewing the previous works. To determine the best classification approach for envisaging heart disease, we examined the performance of the classification algorithms.

3. Planned System

The initial stage in the planned system (shown in fig 1) is to gather data, which is typically obtained via medical records and test reports. It collects patient data as well as wellness data that may be seen from afar and sent to the next generation. The process of obtaining information from various data which is only exposed implicitly is known as material procurement.

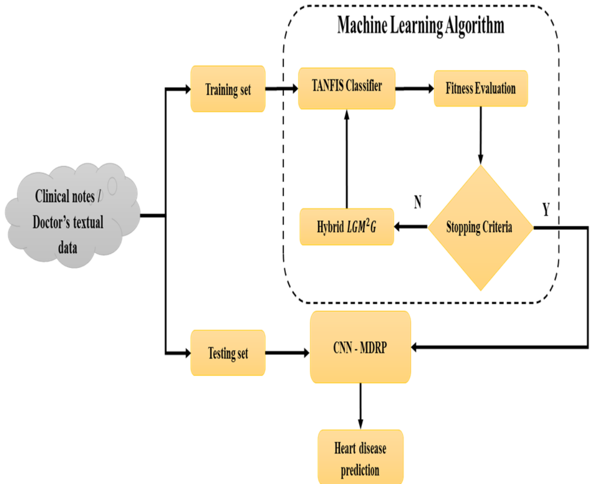

Fig 1. Heart Disease prediction model

3.1 *Data Acquisition and processing*

Data acquisition is made up of two words: data and acquisition. Data refers to raw facts and numbers that can be structured or unstructured, and acquisition refers to gathering data for a specific goal. The organised data comprises both test results and basic patient information such as age, gender, and lifestyle habits. We work with structured data by consulting hospital professionals to identify valuable aspects. The characteristics for unstructured text data are chosen automatically using a CNN algorithm. Before data can be saved, cleansed, preprocessed, and used for other operations, it must first be acquired from appropriate sources. It is the process of extracting important medical data, converting it into the proper format, and feeding it into a classification and prediction system.
A numerical representation of a word, usually vectors, is called a word embedding. In other terms, a word embedding is a vector that reflects a word's characteristics. The frequently used terminology relevant to conditions such as heart disease would be included in the dataset of our proposed system to avoid utilizing repeated phrases for a simple sentence. This simplifies the diagnosis and prediction of symptoms for us.

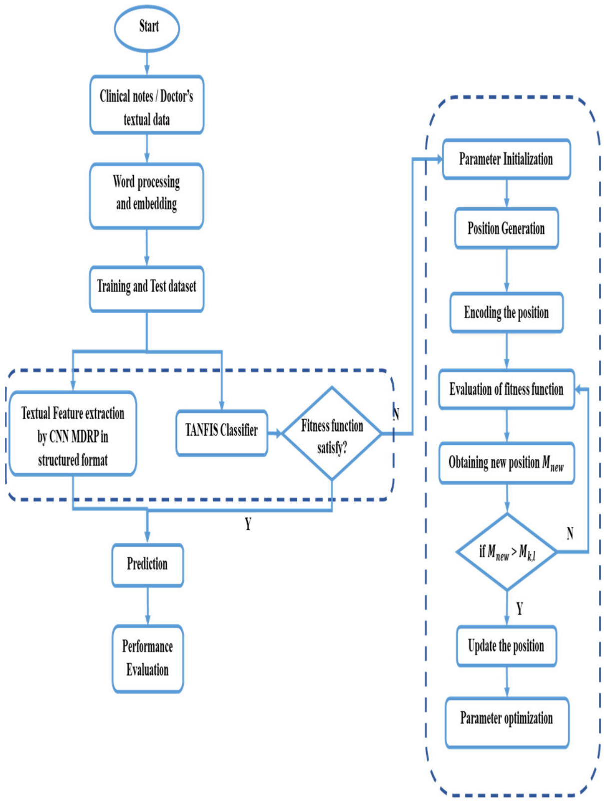

Fig 2: Flowchart of CNN- MDRP based TANFIS classification prediction model

3.2 *CNN-MDRP Risk prediction*

After particular parameter training, a CNN-based uni-modal illness risk detection method is utilised to extract features from the processed data. Based on characteristics in the medical record, the approach determines if an individual is experiencing critical signs of heart disease. As (x1,x2,x3,...,xn), the input contains crucial elements such as gender, indicator factor, and other information. H_0-high risk and H_1-minimal risk are two additional attributes that measure the severity of a risk factor. Textual data seems to be an unorganized assemblage of data. It is difficult to identify crucial elements for prediction in this situation. The risk factor of an specific reported in the dataset is calculated using both organised and disorganized data in a multidimensional database. Mutually structured and disorganized data are lacking for analysing illness forecast in general. Because the above-mentioned method takes time, we employ the CNN-MDRP algorithm to solve this problem.

3.3 *TANFIS Classifier*

To predict if a patient has heart disease, the TANFIS classifier is used to create parameters from training data. The classifier incorporates a customized method to enhance education and edition parameters. This uses artificial intelligence to pretend to be human in order to solve complicated tasks. The data learning technique use fuzzy inference to turn neural network element inputs into desired outputs, while adaptive learning parameters are used to maximise the fuzzy inference system's membership functions in order to train the data. Fuzzification, product, normalisation, de-fuzzification, and summation are the five layers of the TANFIS classifier. The network's efficiency is totally dependent on the nodes' changeable parameters. In the TANFIS classifier's fuzzy inference method, two rules are taken into account.

Rule 1: if A is X_1 and B is Y_1,
then $o_1 = a_1A + b_1B + c_1$, (1)
Rule 2: if A is X_2 and B is Y_2,
then $o_2 = a_2A + b_2B + c_2$, (2)

For A and B input vectors, X_1, X_2, Y_1, and Y_2 are fuzzy inference sets. The de-fuzzification level-4 parameters are a1, a_1, a_2, b_1, b_2, c_1, and c_2. The TANFIS framework's output can be represented as "o".

The following are the functions of the various levels of the TANFIS classifier:

Level-1: In the level-l, the nodes are considered adaptive nodes. Every node at this layer has a degree of membership that, based on the corresponding fuzzy set, generates a new membership degree. As a result, this level achieves fuzzification. The generalised bell-shaped membership function is represented by the output vector. The Gaussian membership function is defined as follows:

$$P_K^1 = \mu X_k(a) \quad for \quad k = 1,2 \tag{3}$$
(3)
$$P_K^1 = \mu Y_{k-2}(b) \quad for \quad k = 3,4 \tag{4}$$
(4)
$$\mu X_k(a),\ \mu Y_{k-2}(b) = exp\left(-\left(\frac{x-b}{\sigma}\right)^2\right) \tag{5}$$

The square node with node function is represented by k. The premise parameters are denoted by the parameters σ and b. For input vectors A and B, X_k and Y_k denote linguistic labels. Furthermore, μX_k and μY_k denote membership functions that are in the range [0, 1]. In Eq 5, the estimate of μX_k and μY_k is shown.

Level-2: The output is evaluated using t-norm operators as from incoming signal in level 2, and the vertices are generally fixed nodes. The outcome of each node reaches the rule's ring strength. The result of level-2 is as continues to follow:

$$P_K^2 = z_k = \mu X_k(a).\ \mu Y_{k-2}(b) \quad for \quad k = 1,2 \tag{6}$$

Level-3: It has fixed nodes and so executes the normalisation (N) procedure. Every node standardizes the previous level's node purpose. It also denotes the ratio of a rule's ring intensity to the sum of all of the rule's ring strengths. A rule's normalised ring strength can be calculated as follows:

$$P_K^3 = \overline{z_k} = \frac{z_k}{z_1 + z_2} \quad for\ k = 1,2 \tag{7}$$

Level-4: The adaptive nodes are outlined at level 4, which represents the defuzzification level. It creates a flexible relationship between the standardized ring intensity as well as the component value that results. As demonstrated in Eq. 8, the output may be obtained by multiplying level-3 and 1values.

$$P_K^4 = \overline{z_k}.g_k$$

$$P_K^4 = \overline{z_k}.g_k = z_k.(\ a_kA + b_kB + c_k) \tag{8}$$

where a_k, b_k, c_k signifies constraint set.

Level-5: Summation () are the plane's endpoints, and they impact the accurate result by summing all acknowledged data. The arriving indications from the previous level are combined, and the outputs are converted to brittle values.

$$P_K^5 = \sum \overline{z_k}.g_k = \sum \frac{z_k g}{z_k} \tag{9}$$

The TANFIS classifier's rules and membership function may be learned using the greatest quantity of recognized information relatively than a sample of anticipated data. The back propagation technique was used to optimise the premise parameters, while the least square mean was used to change the subsequent parameters. Figure 2 shows the parameter optimization of the TANFIS classifier model. A hybrid LGM^{2}G algorithm was employed to accomplish ideal tuning in the TANFIS classifier's membership function. The goal of using the hybrid LGM^{2}G method is to improve the TANFIS classifier parameters by transferring optimum weights to levels 4 and 5. The LGM^{2}G method obtains the predicators first, and then divides the data into training and testing datasets. Following that, the best parameters are input into the TANFIS classifier to train it using Eq (10)

$$\text{Mean square error (MSE)} = \sum_{k=1}^{N} \frac{(x_k - y_k)^2}{N} \tag{10}$$

where x_k - genuine value, y_k - target productivity value, and N - size of contributed samples.

The fitness matrix is usually evaluated using the MSE. It is used to approximate the regression performance results without affecting the algorithm's computing cost. The best solution is determined by the smallest difference between definite and forecasted output values.

Finally, the best solution is supplied into the data classification and prediction process.

4. Evaluation Results

True positive (TP- the quantity of cases appropriately correctly classified as considered necessary), false positive (FP- the quantity of cases inaccurately foreseen as considered necessary), true negative (TN- the amount of cases properly classified as not needed), and false negative (FN- the quantity of cases inaccurately indicated as not desirable) are the terms used in the measuring performance. Then, as follows [14][16], we may obtain four measurements:
Accuracy: It is defined as the number of variables in the dataset that have been effectively categorised.

$$\text{Accuracy} = (TP + TN)/ (TP + TN + FP + FN)$$

Precision: It is defined as the proportion of correctly categorised things to ones that are incorrectly categorised.

$$\text{Precision} = TP / (TP + FP)$$

Recall: Over all the positive examples in the data, recall is a statistic of how many of the positive cases the classifier correctly predicted. Sensitivity is a term that is occasionally used to describe it.

$$\text{Recall} = TP / (TP + FN)$$

F1-score: It's defined as the number of objects that have been categorised erroneously.

$$\text{F1-score} = 2 * TP/ (2 * TP + FP + FN)$$

When compared to other approaches of existing methodologies, the following table provides a clear picture of reduced complexity and the best accuracy rate.

TABLE 1: Table of accuracy rates based on different approaches

J48	84%
FFT	81%
Fuzzy Weighted AIRS	81%
SVM with Gaussian Bayes	83%
SVM	88.30%
PSO-TSVM	90%
ISAGSO-WFSVM	92%
Neuro-Genetic approach with CNN-MDRP	96.25%
CNN-MDRP based TANFIS	98.40%

Fig 1 depicts the suggested CNN-MDRP based TANFIS approach's training model. Integrate LGM^{2}G algorithm for feature optimization in CNN-MDRP to reduce the nature of complexity.

Comparison graph of accuracies with the existing system is given in fig 3.

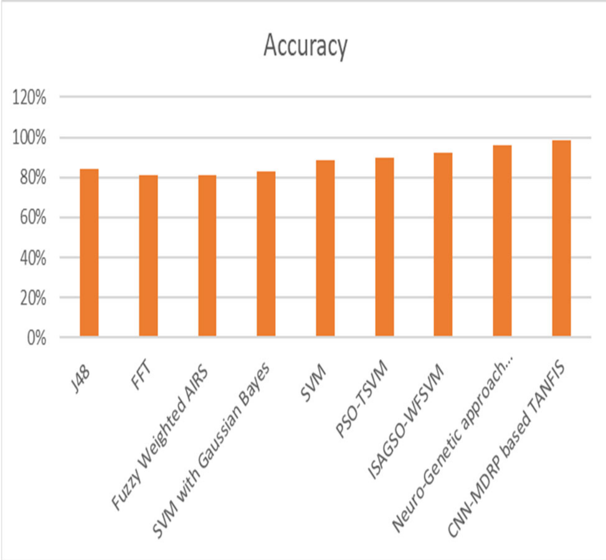

Fig. 3 Comparison of accuracies with existing system

The suggested study uses a CNN-MDRP inspired TANFIS classifier to accurately forecast cardiac disorders. It divides the input data into two categories: heart disease and no heart disease (see Fig 4).

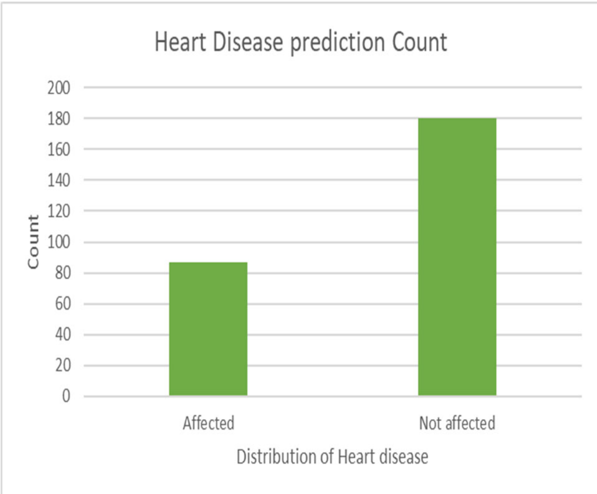

Fig. 4 Classification results of prediction of heart disease

5. Conclusion

The combination of CNN-MDRP with the TANFIS classifier improves the accuracy of chronic heart disease identification. The duty of preprocessing is carried out by CNN-MDRP, which provides its best in illness prediction via multiple convolutional hidden layers. The LGM2G method balances both exploitation and exploration characteristics in the entire search space during the training phase, initiating the optimization approach to get the optimal values for TANFIS classifier parameters. Various simulated circumstances and datasets are used to analyse the performance findings. The proposed method's results are compared to those of many optimization approaches, and the results demonstrate that it achieves 98.40 percent accuracy. The accuracy findings, in particular, have demonstrated the suggested method's robustness and stability. As a result, the suggested hybrid strategy would be effective in improving a health monitoring system for accurate classification and prediction.

References

[1] V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja “Heart disease prediction using machine learning tech : A survey” International Journal of Engineering & Technology, 7 (2.8), April 2018.

[2] T.Nagamani, S.Logeswari, B.Gomathy,” Heart Disease Prediction using Data Mining with Mapreduce Algorithm”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.

[3] Avinash Golande, Pavan Kumar T, ”Heart Disease Prediction Using Effective Machine Learning Techniques”, International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019.

[4] Theresa Princy R,J. Thomas,’Human heart Disease Prediction System using Data Mining Techniques’, International Conference on Circuit Power and Computing Technologies,Bangalore,2016.

[5] W. Dai, T. S. Brisimi, W. G. Adams, T. Mela, V. Saligrama, and I. C. Paschalidis, “Prediction of hospitalization due to heart diseases by supervised learning methods”, International Journal of Medical Informatics, Vol.84, No.3, pp.189–197, 2015

[6] Vinitha S, Sweetlin S, Vinusha H, Sajini S. “Disease Prediction Using Machine Learning Over Big Data”. Computer Science & Engineering: An International Journal (CSEIJ), Vol.8, No.1, [2018]. DOI:10.5121/cseij.2018.8101.

[7] Sayali Ambekar and Dr.Rashmi Phalnikar. “Disease Prediction by using Machine Learning”. International journal of computer engineering and applications, Volume XII, special issue, May 18. ISSN: 2321-3469.

[8] A. F. Otoom, E. E. Abdallah, Y. Kilani, A. Kefaye, and M. Ashour,"Effective diagnosis and monitoring of heart disease", International Journal of Software Engineering and Its Applications, Vol.9, No.1, pp. 143-156, 2015.

[9] V. Chaurasia and S. Pal, “Data mining approach to detect heart diseases”, International Journal of Advanced Computer Science and Information Technology, Vol.2, No.4, pp.56-66, 2014..

[10] Nagaraj M Lutimath,Chethan C,Basavaraj S Pol.,’Prediction Of Heart Disease using Machine Learning’, International journal Of Recent Technology and Engineering,8,(2S10), pp 474-477, 2019.

[11] Vidhya K, Shanmugalakshmi R. Modified adaptive neuro-fuzzy inference system (M-ANFIS) based multi-disease analysis of healthcare big data. J Supercomput. 2020;76:8657-8678.

[12] Raj S. An efficient IoT-based platform for remote real-time cardiac activity monitoring. IEEE Trans Consum Electron. 2020;66:106-114.

[13] Bharathi R, Abirami T, Dhanasekaran S. Energy efficient clustering with disease diagnosis model for IoT based sustainable healthcare systems. Sustain Comput Inform Syst. 2020;28(100453):1-28.

[14] Khan MA, Algarni F. A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS. IEEE Access. 2020;8:122259-122269.

[15] Khan MA. An IoT framework for heart disease prediction based on MDCNN classifier. IEEE Access. 2020;8:34717-34727.

[16] Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access. 2019;7:81542-81554.

[17] Al-Makhadmeh Z, Tolba A. Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: a classification approach. Measurement. 2019;147(106815):1-15.

[18] Nalluri M, Manisha M, Roy D. Hybrid disease diagnosis using multiobjective optimization with evolutionary parameter optimization. J Healthcare Eng. 2017;5907264:1-27.

[19] S.-H. Wang, T.-M. Zhan, Y. Chen, Y. Zhang, M. Yang, H.-M. Lu, H.- N. Wang, B. Liu, and P. Phillips, “Multiple sclerosis detection based on biorthogonal wavelet transform, rbf kernel principal component analysis, and logistic regression,” IEEE Access, vol. 4, pp. 7567–7576, 2016.

[20] S. Bandyopadhyay, J. Wolfson, D. M. Vock, G. Vazquez-Benitez, G. Adomavicius, M. Elidrisi, P. E. Johnson, and P. J. O’Connor, “Data mining for censored time-to-event data: a bayesian network model for predicting cardiovascular risk from electronic health record data,” Data Mining and Knowledge Discovery, vol. 29, no. 4, pp. 1033–1069, 2015.

[21] J. Wan, S. Tang, D. Li, S. Wang, C. Liu, H. Abbas and A. Vasilakos, “A Manufacturing Big Data Solution for Active Preventive Maintenance”, IEEE Transactions on Industrial Informatics, 2017.

[22] W. Yin and H. Sch¨utze, “Convolutional neural network for paraphrase identification.” in HLT-NAACL, 2015

[23] S.-M. Chu, W.-T. Shih, Y.-H. Yang, P.-C. Chen, and Y.-H. Chu, “Use of traditional chinese medicine in patients with hyperlipidemia: A population-based study in taiwan,” Journal of ethnopharmacology, 2015