Coronary Heart Disease Prediction Using GKFCM with RNN

Varun Malik
Chitkara University Institute of Engineering and Technology
Chitkara University
Punjab, India.
varun.malik@chitkara.edu.in

Ruchi Mittal
Chitkara University Institute of Engineering and Technology
Chitkara University
Punjab, India.
ruchi.mittal@chitkara.edu.in

Ajay Rana
Amity University
Greater Noida
Uttar Pradesh, India.
ajay_rana@amity.edu

Irfan Khan
Lloyd Institute of Engineering and Technology
Greater Noida, Uttar Pradesh, India
dean.iqac@liet.in

Pankaj Singh
Lloyd Law College
Greater Noida
Uttar Pradesh, India.
Pankaj.singh@lloydlawcollege.edu.in

Bashar Alam
Lloyd Institute of Engineering and Technology
Greater Noida, Uttar Pradesh, India
shekhar.singh@lloydpharmacy.edu.in

2023 6th International Conference on Contemporary Computing and Informatics (IC3I) | 979-8-3503-0448-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/IC3I59117.2023.10398020

***Abstract*— For humans, coronary heart disease is the number one killer. A coronary angiogram is the most common method for identifying coronary heart disease. The patient risks experiencing a bodily injury, problems, and bad reactions during this procedure. Coronary angiography has a hefty price tag. However, the heart color Doppler echocardiography report, biochemical blood markers, and other essential data may reveal the amount of heart damage in a patient. This article proposed for predicting coronary heart disease using Gaussian Fuzzy C-Means Clustering (GKFCM) with Recurrent Neural Network (RNN). This study predicts coronary heart disease based on datasets. The datasets are normalized then the dataset has clustered using GKFCM and heart disease prediction using hybrid method as GKFCM with RNN. The Experimental results show the proposed model has achieved the highest accuracy at 99%.**

Keywords—Coronary heart disease, GKFCM, Prediction, RNN, Clustering, Classification

I. INTRODUCTION

Worldwide, coronary heart disease is the "primary killer of human health," The World Health Organization WHO estimates that in 2019, cardiovascular disease will account for 31% of all deaths worldwide [1]. As a result, it is critical to develop a low-cost technology for quickly and adequately diagnosing coronary heart disease [2].

The most common method for identifying coronary heart disease is coronary angiography [3]. However, there are several risks for patients associated with coronary angiography: Start an irregular heartbeat. (2) Contrast agent-related adverse effects, including allergic reactions and kidney damage [4]. Third, there is the vagus nerve reflex. The inspection procedure is rigorous and time-consuming [5]. At the puncture site, bleeding, a hematoma, a pseudo aneurysm, and an arteriovenous fistula will emerge. Occasionally, it takes time to halt the bleeding once therapy has been completed. [6-9]. Five patients requiring femoral artery puncture will need three days of hospitalization. Coronary angiography uses radiation, which has adverse effects on humans. People with modest incomes may have trouble affording coronary angiography because of its high price [10–14].

Biochemical blood test markers will disclose the extent to which coronary arteries are blocked. Myocardial infarction causes the release of cardiac troponin complexes into the circulation, where they stay for a very long time and progressively grow in concentration [15]. Thus, the blood troponin level is often evaluated in patients who exhibit signs of cardiac illness [16]. Doctors might use the test's findings to diagnose heart problems. So, it's possible that blood tests might be used to predict cardiovascular disease's danger [17]. In addition, coronary heart disease may be diagnosed based on a heart colour Doppler echocardiogram results. Coronary atherosclerotic heart disease is a kind of atherosclerotic coronary artery lesion that leads to ischemia, hypoxia, and necrosis of the heart muscle [18-20]. This study proposes the GKFCM with RNN model for evaluating coronary artery occlusion severity by integrating information from 10 different attributes, tobacco, alcohol, age and other fundamental patient characteristics.

The model consists of four individual components. The first step is a clustering phase using GKFCM. This problem is solved by using an asynchronous approach. Through the use of training data, we optimize recurrent neural networks with GKFCM used to parameterize the stochastic approach. The main contribution of the paper is combining GKFCM with RNN algorithm.

The remainder of the paper is as follows; section II discusses various existing author's methods. Section III presents the proposed model, section IV discusses results and discussion, and Section V deals with a conclusion and future scope.

II. BACKGROUND STUDY

Anandajayam et al. [1] These authors proposed a heart disease prediction scheme system that can store a vast

quantity of patient data and then be utilized to forecast coronary heart disease in a patient. Consequently, the patient may take the required precautions to prevent disease development. This treatment enables the patient to build awareness to prevent coronary heart disease.

Chotwani, P. et al. [3] reviewed the Cleveland dataset, and the decision tree approach was used to provide findings. It will aid in understanding and forecasting heart disease risk factors, allowing patients to be treated quickly and efficiently. The decision tree approach was successfully applied to the given dataset, and the resulting outcomes vary from zero to four. The values were saved in the num attribute (nominal). It may be utilized efficiently in a variety of biological studies. The dataset contains a plethora of risk factors for cardiovascular disease. Large hospitals sometimes neglect or overlook these components.

Mahboob, T. et al. [9] numerous heart diseases were the most costly and pervasive health issue plaguing today's society. As a result, early illness identification helps avoid future suffering by applying simple lifestyle changes. The primary objective was to discover the most promising strategy by analyzing all disseminated technologies.

Rahman, M. et al. [11] Heart disease was growing increasingly common worldwide. The issue was now showing up in highly young people. This led to a rise in erosion rates. This is an emergency that has to be fixed right now. Despite a plethora of data in the healthcare sector, there was a lack of understanding. As a result, the author must unearth this knowledge to enhance treatment. Data mining methods extract critical information from vast datasets for diagnostic research.

Shanmugasundaram, G. et al. [13] these authors research examines and report on the critical factors of heart disease and research papers on heart disease prediction. It has been discovered that all researchers considered not all attributes. Few studies remove a few characteristics to increase accuracy. The author examined the outstanding concerns raised by various research attempts to predict heart disease.

Sultan Bin Habib, A.-Z. et al. [15] These authors' research examines many boosting-based ensemble machine learning (ML) algorithms applied to the informative index, including AdaBoost, GBM, XGBoost, LGBM, and boost. It looks at factors including blood pressure, cholesterol, and diabetes to forecast the chance of coronary disease in the following ten years. This research will help identify people who may develop coronary disease during the following decade. This might help establish preventive measures and protect the patient from coronary disease. As a result, when a patient was suspected of having coronary disease, clinicians may thoroughly examine the patient's treatment information.

Yekkala, I. et al. [20] Predicting heart disease using particle swarm optimization (PSO) and an ensemble classifier. These authors investigate PSO as a feature selection approach to exclude the lowest-ranked characteristics. Then, ensemble techniques were used as a classifier to lessen the number of false positives and boost the accuracy of the classification.

III. Materials And Methods

The coronary heart disease prediction has been made with GKFCM and RNN techniques.

A. ***Dataset Gathering***: The datasets are gathered from https://www.kaggle.com/datasets/billbasener/coronary-heart-disease. The dataset contains 10 columns with 463 data records and a 21KB memory size.

B. Clustering Using Gaussian kernel fuzzy c-means clustering (GKFCM)

The partition matrix U is first set up by fixing c, where c is (2=cn), and choosing a value for the parameter m (0). In this method, each operation will be represented by a letter, r, where r=0, 1, 2, We are to calculate each step's c centre vector {Vij}.

$$V_{ij} = \frac{\sum_{k=1}^{n}(\mu ik)m\ xkj}{\sum_{k=1}^{n}(\mu ij)m} \quad (1)$$

Calculate the distance matrix

$$D_{ij} = \sum_{j=1}^{m}(xkj - vlj)2 \quad (2)$$

C. Heart disease prediction using GKFCM with Recurrent Neural Network (RNN)

Update the partition matrix for the r[th] step $U^{(R)}$ as

$$\mu^{r-1}ij = (1/\sum_{j-1}^{c} dik/djk)^{\frac{2}{m}-1}) \quad (3)$$

IF || U(k+1) – U(k)||<δ, then we are to stop. Without iteratively revising the cluster centres and the membership grades for data points, we must return to step 2.

A dataset's cluster centers may be moved to the right spot with the help of FCM. The Fuzzy C-Means technique is a gateway towards the more advanced Fuzzy logic used in K-Means. In particular, FCM clustering algorithms often produce a cluster whose membership weights may be intuitively understood but are improbable.

Medical disease prediction is only one field where the RNN classifier has found widespread usage. Because it is a dynamic system that organizes the information in the input sequence, the RNN classifier is a more effective tool for data categorization and prediction. Learning an approximate identity function from a collection of unlabeled training data is an example of an Auto Encoder (AE) approach. Typical AE models include a single hidden layer and use dimensionality reduction to get their results.

The encoding process takes input data represented as xR(d x) and maps the hidden layers according to the equation,

$$y = \sigma(Wx + b) \qquad (4)$$

In the input layer, the weight matrix is represented as$W \in R^{d_y d_x}$, the bias vector is represented as$b \in R^{d_y}$, and the non-linear activation function is given as. So, the equation looks like this for the final goal: (5),

$$Z = \sigma'(W'y + b') \qquad (5)$$

In contrast, the output layer's weight matrix and bias vector are written as $W'y + b'$ and $b \in R^{dy}$ and$b' \in R^{dy}$, respectively. Estimates of the parameters = $W'y$ in the AE training process to minimize the total reconstruction cost for each training set. An equation illustrating the cost breakdown is shown below (6).

$$j = \sum_{x \in d_x} L(x, z) + \mu \sum_{i,j} W_{i,j}\, 2 \qquad (6)$$

Where,

L(x,z) - Estimation of square error

D_x -Training samples in the dataset

μ - Hyper parameter of regularization strength

The RNN method can control how the network's internal states handle a given sequence of input and output values. Similarly, internal states act as a memory by storing data gleaned from external stimuli. The RNN method adopts a single fixed output vector after mapping a succession of fixed input vectors. The schematic shows the general structure of the RNN.

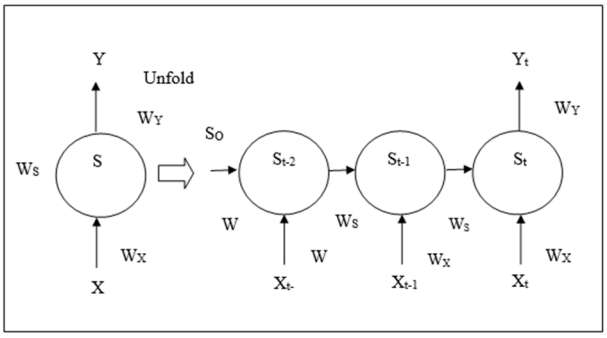

Figure 1 General Architecture of RNN

Take it that is the sequence's input vector and that S is the internal state (which contains the previous state value). Therefore, St is a variable representing the time-coded as a secret state. This equation represents the St. (7),

$$S_t = \sigma(W_x x_t + W_s S_{t-1}) \qquad (7)$$

In contrast, weight matrices are represented by W_x, while Xt denotes the Ws input vector. The output is denoted by Y_t and computed using the following formula: (8).

$$yt = \sigma'(W_y S_t) \qquad (8)$$

The activation function at the output, denoted by ', is weighted by the matrix W y. Preceding equation does not include the bias terms (12). GKFCM provides key disease-related features to the AE-based RNN algorithm. The RNN classifier may reliably predict many diseases and conditions. The dimensionality problem of features is reduced to a minimum using this method the section below, we use appropriate assessment measures to assess the suggested GKFCM+RNN's performance.

IV. Results And Discussion

The proposed model has implemented by using python programming language

The attributes are plotted using scatter is represented in figure 2.

The probabilities for each of the components are shown in Figure 3. The axis labelled "factors" is located on the left, while the axis labelled "probability" is on the right.

The accuracy has been achieved with 99% as the proposed method while comparing with an existing algorithm like ML and DL algorithm has achieved 85 to 95% only.

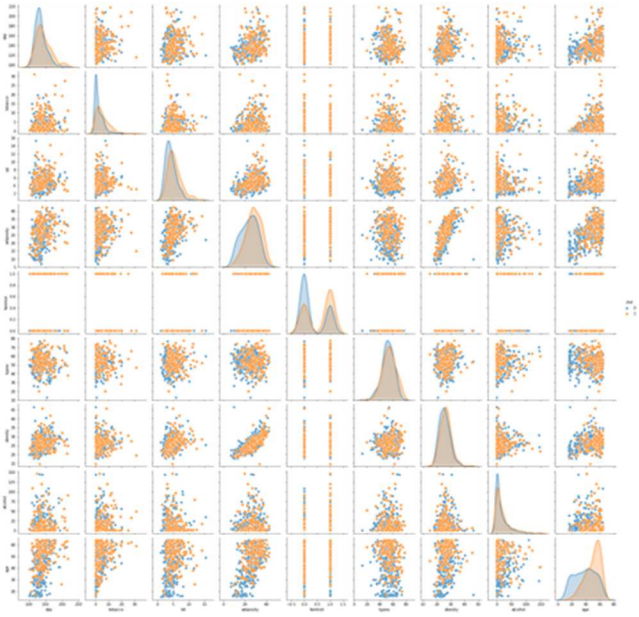

Figure 2: Scatter plot for coronary heart disease

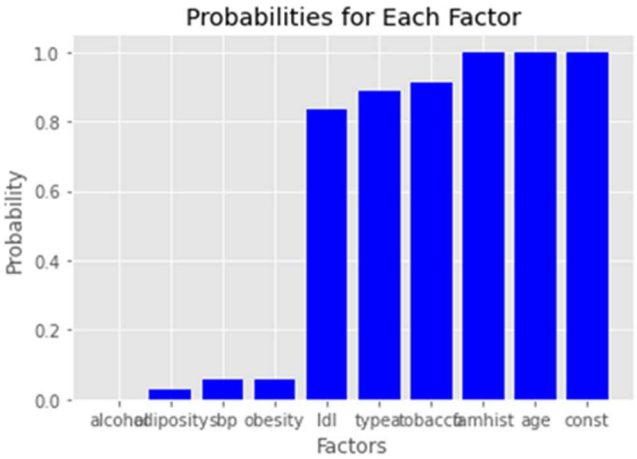

Figure 3: Probabilities for each factor

TABLE I: CLASSIFICATION METRICS

Algorithm	Class	Accuracy	Precision	Recall	F1 Measure
Proposed	0	99	100	98	99
	1	99	98	100	99
	Macro Avg	99	99	99	99
	Weighted Avg	99	99	99	99
DT	Overall	85	87	82	81
RF	Overall	87	88	91	91
CNN	Overall	95	92	93	91
LSTM	Overall	92	90	91	90

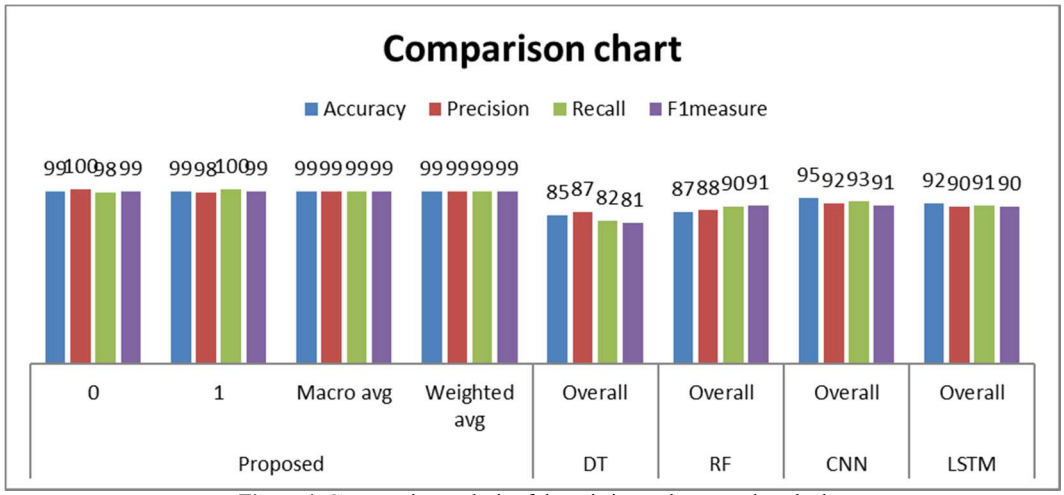

Figure 4: Comparative analysis of the existing and proposed method

V. CONCLUSION

This article proposed coronary heart disease prediction using GKFCM with RNN algorithm. As a pre-requisite to the initial RNN system, a GKFCM classification algorithm was implemented. Based on the results findings, the suggested solution outperforms the hybrid techniques in terms of heart disease types, particularly for long estimate coronary heart disease. The categorized accuracy is 99% which is sufficiently high as compared to other approaches [21][22]. The proposed method can detect heart disease, and the most effective results are reached regarding the precision of coronary heart disease. For further to implement the combine with various disease dataset to predict the disease.

REFERENCES

[1] Anandajayam, P., Krishnakoumar, C., Vikneshvaran, S., & Suryanaraynan, B. (2019). Coronary Heart Disease Predictive Decision Scheme Using Big Data and RNN. 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). doi:10.1109/icscan.2019.8878765

[2] Chauhan, A., Jain, A., Sharma, P., & Deep, V. (2018). Heart Disease Prediction using Evolutionary Rule Learning. 2018 4th International Conference on Computational Intelligence & Communication Technology (CICT). doi:10.1109/ciact.2018.8480271

[3] Chotwani, P., Tiwari, A., Deep, V., & Sharma, P. (2018). Heart Disease Prediction System using CART-C. 2018 International Conference on Computer Communication and Informatics (ICCCI). doi:10.1109/iccci.2018.8441272

[4] Jabbar, M. A., & Samreen, S. (2016). Heart disease prediction system based on hidden naïve bayes classifier. 2016 International Conference on Circuits, Controls, Communications and Computing (I4C). doi:10.1109/cimca.2016.8053261

[5] Krishnani, D., Kumari, A., Dewangan, A., Singh, A., & Naik, N. S. (2019). Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms. TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON). doi:10.1109/tencon.2019.8929434

[6] Li, H., Ren, G., Yu, X., Wang, D., & Wu, S. (2020). Discrimination of the Diastolic Murmurs in Coronary Heart Disease and in Valvular Disease. IEEE Access, 8, 160407–160413. doi:10.1109/access.2020.3021093

[7] Li, W., Chen, D., & Le, J. (2020). Coronary Heart Disease Prediction Based on Combined Reinforcement Multitask Progressive Networks. 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). doi:10.1109/bibm49941.2020.9313275

[8] Lin, S., Zhuang, X., Huang, S., Liu, Y., Shen, L., & Liao, X. (2019). Face Analysis for Coronary Heart Disease Diagnosis. 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). doi:10.1109/cisp-bmei48845.2019.8966020

[9] Mahboob, T., Irfan, R., & Ghaffar, B. (2017). Evaluating ensemble prediction of coronary heart disease using receiver operating characteristics. 2017 Internet Technologies and Applications (ITA). doi:10.1109/itecha.2017.8101920

[10] Mohawish, A., Rathi, R., Abhishek, V., Lauritzen, T., & Padman, R. (2015). Predicting Coronary Heart Disease risk using health risk assessment data. 2015 17th International Conference on E-Health Networking, Application & Services (HealthCom). doi:10.1109/healthcom.2015.7454479

[11] Rahman, M., Zahin, M. M., & Islam, L. (2019). Effective Prediction On Heart Disease: Anticipating Heart Disease Using Data Mining Techniques. 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT). doi:10.1109/icssit46314.2019.8987776

[12] Rane, A. L. (2018). A survey on Intelligent Data Mining Techniques used in Heart Disease Prediction. 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS). doi:10.1109/csitss.2018.8768735

[13] Shanmugasundaram, G., Selvam, V. M., Saravanan, R., & Balaji, S. (2018). An Investigation of Heart Disease Prediction Techniques. 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA). doi:10.1109/icscan.2018.8541165

[14] Sharma, V., Rasool, A., & Hajela, G. (2020). Prediction of Heart disease using DNN. 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). doi:10.1109/icirca48905.2020.9182991

[15] Sultan Bin Habib, A.-Z., Tasnim, T., & Billah, M. M. (2019). A Study on Coronary Disease Prediction Using Boosting-based Ensemble Machine Learning Approaches. 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET). doi:10.1109/iciet48527.2019.9290600

[16] Suvarna, C., Sali, A., & Salmani, S. (2017). Efficient heart disease prediction system using optimization technique. 2017 International Conference on Computing Methodologies and Communication (ICCMC). doi:10.1109/iccmc.2017.8282712

[17] Tithi, S. R., Aktar, A., Aleem, F., & Chakrabarty, A. (2019). ECG data analysis and heart disease prediction using machine learning algorithms. 2019 IEEE Region 10 Symposium (TENSYMP). doi:10.1109/tensymp46218.2019.8971374

[18] Wang, J., Liu, C., Li, L., Li, W., Yao, L., Li, H., & Zhang, H. (2020). A stacking based model for non-invasive detection of coronary heart disease. IEEE Access, 1–1. doi:10.1109/access.2020.2975377

[19] Widiyaningtyas, T., Zaeni, I. A. E., & Wahyuningrum, P. Y. (2019). Self-Organizing Map (SOM) For Diagnosis Coronary Heart Disease. 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). doi:10.1109/icitisee48480.2019.9003746

[20] Yekkala, I., Dixit, S., & Jabbar, M. A. (2017). Prediction of heart disease using ensemble learning and Particle Swarm Optimization. 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon). doi:10.1109/smarttechcon.2017.8358460

[21] Balyan, A. K., Ahuja, S., Lilhore, U. K., Sharma, S. K., Manoharan, P., Algarni, A. D., ... & Raahemifar, K. (2022). A hybrid intrusion detection model using ega-pso and improved random forest method. Sensors, 22(16), 5986.

[22] Trivedi, N. K., Gautam, V., Anand, A., Aljahdali, H. M., Villar, S. G., Anand, D., ... & Kadry, S. (2021). Early detection and classification of tomato leaf disease using high-performance deep neural network. *Sensors*, *21*(23), 7987.