Journal of
Information Systems Engineering
and Business Intelligence

Vol.9, No.2, October 2023
Available online at: http://e-journal.unair.ac.id/index.php/JISEBI

Optimizing Cardiovascular Disease Prediction: A Synergistic Approach of Grey Wolf Levenberg Model and Neural Networks

Sheikh Amir Fayaz [1], Majid Zaman [2]*, Sameer Kaul [3], Waseem Jeelani Bakshi [4]

[1)3)] *Department of Computer Sciences, University of Kashmir, Srinagar, India*
[1)]skh.amir88@gmail.com, [3)]kaulsameer78@gmail.com

[2)] *Directorate of IT & SS, University of Kashmir, Srinagar, India*
[2)]zamanmajid@gmail.com

[4)] *Department of Computer Sciences and Engineering, University of Kashmir, Srinagar, India*
[4)]waseembakshi@gmail.com

Abstract

Background: One of the latest issues in predicting cardiovascular disease is the limited performance of current risk prediction models. Although several models have been developed, they often fail to identify a significant proportion of individuals who go on to develop the disease. This highlights the need for more accurate and personalized prediction models.
Objective: This study aims to investigate the effectiveness of the Grey Wolf Levenberg Model and Neural Networks in predicting cardiovascular diseases. The objective is to identify a synergistic approach that can improve the accuracy of predictions. Through this research, the authors seek to contribute to the development of better tools for early detection and prevention of cardiovascular diseases.
Methods: The study used a quantitative approach to develop and validate the GWLM_NARX model for predicting cardiovascular disease risk. The approach involved collecting and analyzing a large dataset of clinical and demographic variables. The performance of the model was then evaluated using various metrics such as accuracy, sensitivity, and specificity.
Results: the study found that the GWLM_NARX model has shown promising results in predicting cardiovascular disease. The model was found to outperform other conventional methods, with an accuracy of over 90%. The synergistic approach of Grey Wolf Levenberg Model and Neural Networks has proved to be effective in predicting cardiovascular disease with high accuracy.
Conclusion: The use of the Grey Wolf Levenberg-Marquardt Neural Network Autoregressive model (GWLM-NARX) in conjunction with traditional learning algorithms, as well as advanced machine learning tools, resulted in a more accurate and effective prediction model for cardiovascular disease. The study demonstrates the potential of machine learning techniques to improve diagnosis and treatment of heart disorders. However, further research is needed to improve the scalability and accuracy of these prediction systems, given the complexity of the data associated with cardiac illness.

Keywords: Cardiovascular data, Clinical data., Decision tree, GWLM-NARX, Linear model functions

Article history: Received 4 May 2023, first decision 20 June 2023, accepted 22 August 2023, available online 28 October 2023

I. INTRODUCTION

Cardiovascular ailments remain the leading contributor to worldwide fatalities, necessitating precise prognostication for efficient medical treatment. The incorporation of machine learning algorithms has showcased their adeptness in predicting illnesses using vast healthcare information [1]. Detecting cardiovascular disease proves challenging due to various risk factors like high blood pressure, irregular heart rate, and cholesterol levels, requiring cautious handling due to its complexity to prevent severe outcomes. Technological advancements, especially in computer-aided outcome systems, have revolutionized medical practices. Machine learning methods have expedited precise disease prediction within the healthcare sector, emphasizing the critical importance of early identification and patient protection [2-3].

In 2015, cardiovascular disease led to more than 17.7 million deaths worldwide [4]. Successful management of cardiac risk demands accurate decision-making and optimal treatment strategies. A Canadian study employed five

* Corresponding author

ISSN 2443-2555 (online) 2598-6333 (print)
doi: http://dx.doi.org/10.20473/jisebi.9.2.119-135

machine learning models to assess 1-month mortality among congestive heart failure patients, while similar research in China and South Korea focused on intra-hospital forecasts for myocardial ischemia patients. Startlingly, one out of every four deaths in the United States is attributed to cardiovascular disease [5]. Over 92.1 million American adults are afflicted by this condition, underscoring the necessity for a highly precise and comprehensive cardiovascular risk prediction system [6-8].

Anticipated advancements in machine learning are poised to reshape clinical practices in the healthcare sector, necessitating the understanding and adoption of these methods by both academics and medical professionals. Despite the availability of risk prediction algorithms, many only consider a limited set of risk factors, thus challenging the effectiveness of risk prediction and classification in intricate scenarios [9]. Coronary heart disease poses the risk of inadequate blood pumping, resulting in symptoms like fatigue, weakness, swollen extremities, and breathlessness. As the healthcare industry generates an abundance of health data, medical records offer insights into the indicators and behaviors associated with cardiovascular disease. A battery of tests and measurements, including blood pressure, cholesterol levels, electrocardiograms (ECGs), and blood sugar assessments, inform diagnostic decisions and treatment strategies [9]. It is important to note that occasional misdiagnoses may occur within healthcare due to human limitations.

The probability of experiencing cardiac arrest has increased during the current situation of halted life. Patients' health can worsen if they delay seeking medical attention for chest discomfort due to fears of contracting a contagious disease. Accurate predictions are crucial for diagnosis and treatment. Researchers [10-11] are continuously developing valuable decision support systems. Diagnosing heart disease remains challenging, with classification methods playing a significant role in prediction. The objective of this research is to introduce a machine learning-driven framework designed to accurately forecast cardiovascular diseases. To attain this goal, a selection of established machine learning algorithms, such as REP Tree, M5P Tree, JRIP, and others, are employed to classify widely acknowledged cardiovascular datasets [12-15]. The success of the chosen classification algorithm in identifying instances of cardiovascular disease is used as a criterion to determine the optimal machine learning algorithm for the task.

The objectives of a study using the GWLM-NARX model for cardiovascular data analysis encompass various goals. Firstly, it aims to assess the likelihood of cardiovascular disease in different patient populations by developing prediction models that incorporate specific patient data such as age, sex, and medical history. Secondly, the model facilitates the improvement of cardiovascular disease treatment plans by enabling researchers to predict patient responses to different therapies. This allows for personalized treatment programs and the optimization of existing ones. Additionally, the GWLM-NARX model aids in the development of novel cardiovascular disease diagnostic tools, as researchers can utilize the model to identify new biomarkers and diagnostic criteria for early recognition and diagnosis of various illnesses. Furthermore, the model contributes to a better understanding of the underlying causes of cardiovascular diseases, as it helps researchers explore the intricate dynamics of the circulatory system and unravel the mechanisms behind disease onset and progression. These discoveries can influence the creation of innovative therapies and interventions.

The use of the GWLM-NARX model for cardiovascular data analysis introduces several motivations and novel aspects. Firstly, this model combines the strengths of the generalized weighted linear model (GWLM) and the nonlinear autoregressive with exogenous input (NARX) model, offering a more comprehensive understanding of the complex dynamics within cardiovascular data. The GWLM captures linear relationships and the influence of various factors on the cardiovascular system, while the NARX model captures nonlinear dynamics and the impact of past inputs on the current output. By leveraging the combined power of these models, the GWLM-NARX approach provides a more accurate and nuanced analysis of cardiovascular dynamics.

Secondly, the application of the GWLM-NARX model holds significant potential for advancing the development of effective and personalized treatment strategies for cardiovascular diseases. Through an accurate representation of the cardiovascular system, the model enables researchers to gain insights into the underlying mechanisms of diseases and identify potential intervention targets. Additionally, the model can predict individual patient responses to different treatments, facilitating the optimization of treatment plans based on specific patient characteristics. The novelty of this approach lies in its integration of the GWLM and NARX models specifically for cardiovascular data analysis, paving the way for new discoveries and advancements in the field of cardiovascular research.

The paper's general format adheres to the basic format, and section 2 of that format contains a review of the literature. The dataset utilized in this investigation is described in Section 3 along with the methods used. The architecture and prediction model utilized in this study are defined in Section 4. Section 5 then describes the experimental analysis, which includes database description, experimental findings, evaluation, performance metrics, and comparison analysis. The examination of the model performance has been expanded upon in section 6, and section 7 summarizes the article with some recommendations for the future.

II. Literature Review

This section has examined and evaluated various machine learning strategies, enabling the creation of predictive models for heart disease. One of the key benefits applicable to a range of medical scientific challenges is the diverse array of approaches and methodologies available in data mining. Numerous successful research endeavors have utilized techniques such as clustering, association rule mining, regression, and classification to anticipate cardiovascular conditions. These methods have been employed to improve disease diagnosis with noteworthy precision and minimal error rates. Based on the existing body of research, classification, as compared to alternative methods, holds paramount importance in cardiac problem prediction. The literature suggests that many research undertakings have pursued individual machine learning techniques or ensemble processes to achieve higher accuracy and reliability in estimation techniques.

Rajamhoana et al. [16] investigated the UCI dataset, which had 12 characteristics. The researchers used artificial neural networks to analyze the data, and they were able to predict heart disease with an impressive 85.66% accuracy. In a study conducted by Babu et al. [17], an exploration of numerous risk factors associated with heart patients led to a research endeavor focused on early heart disorder diagnosis. Through the utilization of various learning classifiers on a selected dataset, Support Vector Machine (SVM) emerged as the most effective, boasting an accuracy of 84.33%.

Krittanawong et al. [18] delved into the overall efficacy of machine learning algorithms for predicting cardiovascular diseases in a 2019 publication. They employed the area under the curve metric to assess predictive capabilities, demonstrating promise in foreseeing conditions like coronary artery disease, irregular heart rhythms, heart failure, and stroke. However, due to the diversity inherent in machine learning algorithms, identifying the optimal choice for cardiovascular ailments remains a challenge.

Examining the impact of COVID-19-induced quarantine measures, Lippi et al. [19] scrutinized the potential emergence of cardiovascular issues during the pandemic. Stringent lockdowns and reduced physical activity heightened the risk of cardiovascular complications, as highlighted by the World Health Organization's guidelines. Post-quarantine, detrimental health effects became evident, prompting the authors to advocate for the continuation of physical activity even during quarantine. This concept influenced the present study within the Computational Intelligence and Neuroscience domain.

Aryal et al. [20] proposed a methodology employing machine learning algorithms to detect microbiota-linked cardiovascular diseases. Analysis of ribosomal RNA 16S from fecal samples, sourced from both cardiovascular and non-cardiovascular patients via the American Gut Project, constituted the core of their research. Five primary categories of machine learning techniques, including decision trees, random forests, neural networks, elastic nets, and support vector machines, were employed for training. Distinct bacterial taxa were identified, with the random forest technique exhibiting a superior characteristic curve of 0.70. Leveraging the established proficiency of random forest in predicting cardiovascular issues, this investigation also incorporated another machine learning algorithm.

In a study by Ketut et al. [21], data from Harapan Kita Hospital was employed to predict cardiac disorders using 18 factors. K-Nearest Neighbors (KNN) was utilized, both with and without parameter weighting, resulting in accuracy levels of 75.11% and 74.0%, respectively. Furthermore, Naive Bayes and SVM were applied to the same dataset, although the outcomes from these models did not yield particularly significant results. Randa et al. [22] focused on cardiac valve disease prediction using the Collective Heart Disease dataset (CAD). With 13 characteristics utilized for model training, Naive Bayes classification demonstrated the highest accuracy of 92%.

A Clustered CART Framework was suggested by Shen et al. [23] to handle the numerous identified rules. Ten features were applied during testing and training. The missing values were then substituted with global means after these features underwent preprocessing, which involved discretizing the continuous features. Then, CARS were generated for the prediction of chronic illness using an already-existing ARM technology.

The authors of the cited paper [4] conducted an extensive assessment of diverse machine learning (ML) strategies aimed at creating automated diagnostic systems for detecting heart disease. They examined various types of data modalities, including clinical features, imaging, and electrocardiography (ECG), within their analysis. Covering studies published from 1995 to 2021, the authors critically evaluated and systematically organized prior research endeavors. By doing so, they highlighted shortcomings and deficiencies present in earlier approaches for automated heart disease detection. Furthermore, the authors outlined prospective avenues for future investigation in the realm of ML-based automated heart disease detection, aimed at enhancing the effectiveness of previous methodologies. This comprehensive review stands as a valuable reference for individuals interested in engaging with the field of automated heart disease detection.

In this paper [24], the authors proposed a novel method for detecting heart disease using a combination of electrocardiogram (ECG) signals and clinical features. The proposed method utilized a hybrid feature selection technique that combined recursive feature elimination and principal component analysis. Following this, a hybrid machine learning framework was created, employing a support vector machine (SVM) classifier integrated with a radial basis function kernel for classification purposes. The effectiveness of the proposed model was assessed using a publicly accessible dataset, yielding an impressive accuracy of 95.06%. Notably, this performance surpassed that of other contemporary leading machine learning models. These findings indicate the potential viability of the proposed technique as a promising avenue for precise and effective heart disease detection.

The literature reviewed in this study examines the use of machine learning approaches to predict and diagnose heart disease with notable accuracy and low error rates. The researchers used different machine learning techniques such as clustering, association rule mining, regression, and classification. Categorization, in particular, is important for predicting cardiac problems. The review highlights the limitations and weaknesses of previous studies and discusses future research directions for the field of automated heart disease detection. The paper also considers the impact of COVID-19 quarantine on cardiovascular illness and the need to maintain physical activity even during quarantine.

TABLE 1
OVERALL DESCRIPTION OF THE MODELS USED BY VARIOUS RESEARCHERS (TABULAR REPRESENTATION)

Year of Publication	Parameters used	Dataset	Model	Overall Accuracy (%)	Ref.
2018	16-14	Cleveland	Basic DT model	93.20	[25]
2019	10	General CVD	Simple LR	87.20	[26]
2019	11	CVD Dataset	NB	82.60	[27]
2017	9	Framingham	KNN	66.70	[28]
2019	13-11	UCI Dataset	HRFLM	88.70	[29]
2020	30	CVD (China)	Random Forest, Ada Boost, CART	78.72; 78.62; 70.25	[30]
2021	10-13	Stanford (same data set)	J48	70.77	[31]
2021	10-13	Stanford (same data set)	Tensor flow, Keras, PyTorch	80.00	[13]

Since various researchers in their studies based on the aforementioned literature have adopted several traditional and ensembled methodologies, a generic tabular representation of their accuracies and computations has been presented in the Table 1.

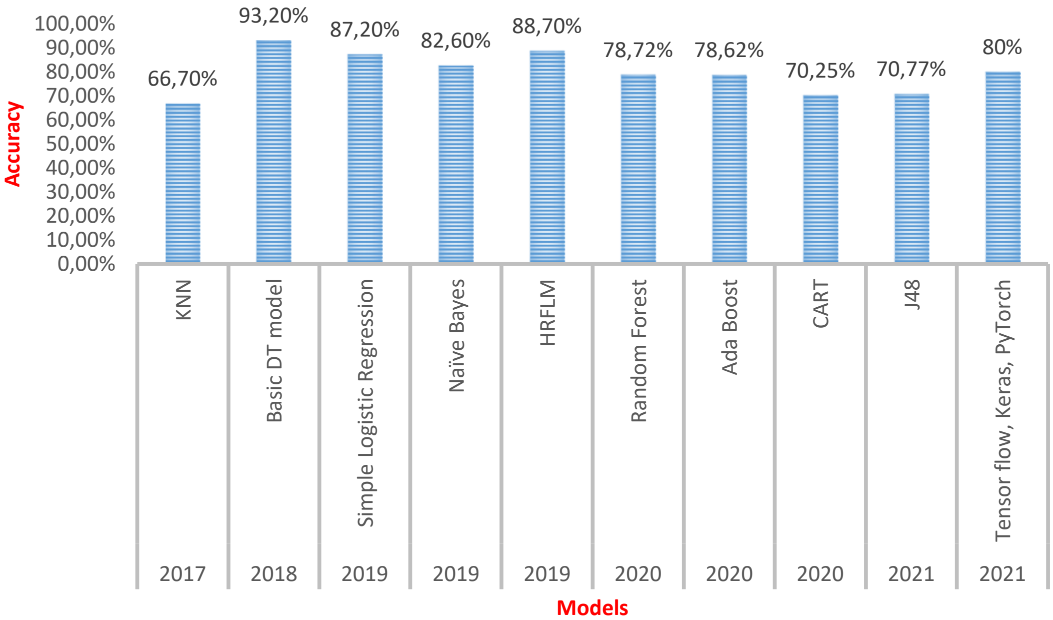

Fig. 1 Overall description of the models used by various researchers (graphical representation)

The graphical representation of different model accuracies proposed by many authors is shown in Fig. 1. In the current scenario of life being on pause, the likelihood of cardiac arrest has heightened. Patients' well-being declines as they delay seeking medical attention for chest discomfort due to their apprehension about contracting an infectious disease. Precise forecasts are imperative for effective diagnosis and treatment. Leading classification algorithms including DT, LMT, M5P Tree, LR, NB, J48, and JRIP have been utilized by numerous studies as written in Table 1 to classify well-known cardiovascular datasets. Researchers are continually developing helpful decision support systems. However, no algorithm has yet emerged as the gold standard for effectively predicting heart disease. There is a significant research gap in the field of clinical sciences, where no research study has been conducted up to this point in the area of cardiovascular prediction employing cutting-edge technology [32]. In this context, research history and effort are still insufficient. Therefore, it is essential to use fresh methodologies in the field of cardiovascular prediction.

According to the literature, there is a lack of widely used methods for predicting cardiovascular diseases using machine learning approaches. The GWLM-NARX model, which combines a generalized weighted linear model (GWLM) with a nonlinear autoregressive model with exogenous inputs (NARX), has shown promising results in predicting cardiovascular diseases. However, it is not yet widely used in the field of medical sciences.

Therefore, by carrying out a study to predict cardiovascular diseases using GWLM-NARX, one could contribute to the development and advancement of AI techniques in the medical field. Additionally, the use of GWLM-NARX could potentially provide a more accurate and efficient approach for predicting cardiovascular diseases, which could lead to better preventative measures and treatments.

Hence, the primary objective of this research is to propose a hybrid machine learning system that enhances the accuracy of cardiovascular disease prediction. In pursuit of this goal, a hybrid model combining GWLM and NARX neural networks has been employed for cardiovascular disease prediction. Consequently, the suitability of the chosen machine learning algorithm is guided by the effectiveness demonstrated by the selected classification approach in instances of cardiovascular ailment [33-34].

III. METHOD

A. Dataset

The primary data set we used to demonstrate the utility of our methodology was collected from the Stanford healthcare repository [35]. The dataset has around 1650 entries, 668 of which are female, and the rest are male. The dataset has 13 parameters, 5 of them are discrete, and 8 are continuous. The dataset contains information from a variety of age groups, with data for females ranging from 15 to 64 and for men from 16 to 64. As a result, it includes participants from three separate generations, including those who first took part, their offspring, and their grandchildren. The dataset is quite trustworthy and produces accurate predictions. The 62 % of samples with N values in the target parameter Cardiovascular disease (CVD) show that the patients do not have any cardiovascular disease. The supplied dataset's properties contain a wide range of data, including demographic and behavioural data as well as details on past and present illnesses. Our suggested framework to predict CVDs uses these characteristics. The dataset's parameters are provided in Table 2.

TABLE 2
PARAMETERS DESCRIPTION USED IN THIS STUDY

No.	Attributes	Values
1.	Gender	15 – 64 (Females), 16 – 64 (Males)
2.	Systolic Blood Pressure (Sbp)	101-218
3.	Tobacco consume level (Cumulative)	Continuous (in Kg's)
4.	Ldl (low density lipoprotein cholesterol level)	Continuous
5.	Adiposity	Continuous
6.	Famhist (Family history disease of heart)	Discrete value (0 – No , 1 – Yes)
7.	Diabetic	Discrete value (0 – No , 1 – Yes)
8.	Smoker	Discrete value (0 – No , 1 – Yes)
9.	Typea (Type A)	Continuous
10.	Obesity	Continuous
11.	Alcohol (Consumption level)	Continuous
12.	Age	Continuous
13.	CHD (coronary heart disease)	Discrete value (0 – No , 1 – Yes)

B. Tools and Techniques for handling the Data: A statistical analysis.

Python is used for exploratory investigation and data processing in this paper. The data is examined in relation to missing values and outliers using the Pandas package. For data visualization and plotting, which aids in data preparation, Seaborn and Matplotlib are utilized. The library of "imblearn" was used to balance the dataset. Sklearn library has made it possible to choose pertinent characteristics and employ classification techniques. High-level mathematical functions can be used with Numpy [13-14] [36]. Using this library, we have calculated the model's accuracy in our research. The overall snapshot of the data used in this study is shown in Table 3.

TABLE 3
SNAPSHOT OF THE CLINICAL DATASET

Gender	sbp	tobacco	ldl	adiposity	famhist	Diabetic	Smoker	typea	obesity	alcohol	age	CHD
M	101	0.48	7.26	13	0	0	0	50	19.82	5.19	16	N
F	102	0.4	3.41	17.22	1	0	0	56	23.59	2.06	39	Y
F	103	0.03	4.21	18.96	0	0	0	48	22.94	2.62	18	N
M	106	5.6	3.2	12.3	0	0	1	49	20.29	0	39	N
F	106	1.61	1.74	12.32	0	1	0	74	20.92	13.37	20	Y
M	106	1.08	4.37	26.08	0	0	1	67	24.07	17.74	28	Y
F	108	0	2.74	11.17	0	1	0	53	22.61	0.95	20	N
M	108	0	1.43	26.26	0	0	0	42	19.38	0	16	N
M	108	0.4	5.91	22.92	1	0	1	57	25.72	72	39	N
M	108	3	1.59	15.23	0	0	1	40	20.09	26.64	55	N
M	108	15	4.91	34.65	0	0	1	41	27.96	14.4	56	N
F	108	1.5	4.33	24.99	0	1	1	66	22.29	21.6	61	Y
M	110	0	7.14	28.28	0	0	1	57	29	0	32	N
F	110	12.16	4.99	28.56	0	0	1	44	27.14	21.6	55	Y
F	110	2.35	3.36	26.72	1	1	1	54	26.08	109.8	58	Y
F	112	0.41	1.88	10.29	0	1	0	39	22.08	20.98	27	N
M	108	15	4.91	34.65	0	0	1	41	27.96	14.4	56	N
F	108	1.5	4.33	24.99	0	1	1	66	22.29	21.6	61	Y
M	110	0	7.14	28.28	0	0	1	57	29	0	32	N

Data visualization is the act of taking raw data and using it to create stunning and elegant graphs, charts, images, and even video that help us understand the statistics and draw conclusions from them. Finding insights from data may be quite challenging, thus data visualization can be utilized extensively to help construct an accurate, reliable pattern and identify trends in the data. The high dimensional multivariate dataset that will be visually evaluated has a correlation between these properties.

The figures below (Figures 2, 3, 4, and 5) depict the relationships between various parameters, such as gender, family history, tobacco use, adiposity, diabetes, smoking, age, lipid profile protein level (LDL), alcohol consumption, obesity, and CHD, based on various plots made using different python libraries, such as Matplotlib, Seaborn, and Ggplot, each of which has its own advantages. In this study, the cardiovascular data, its connection among characteristics and densities, etc., are shown using python libraries.

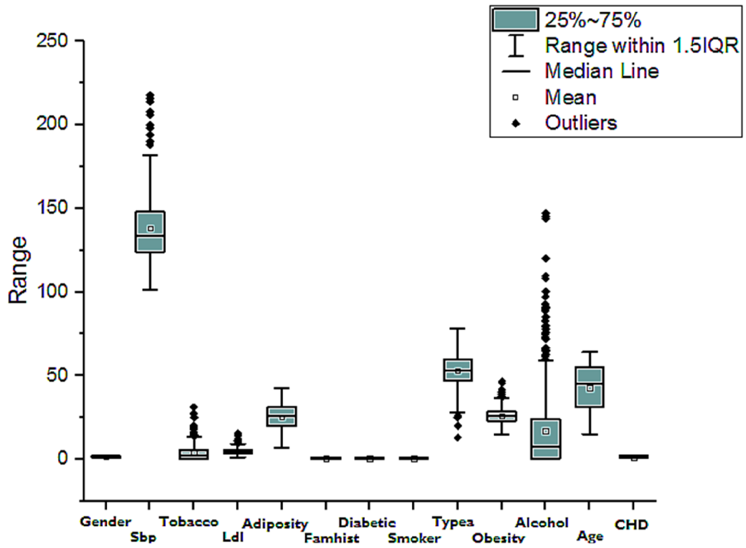

Fig. 2 Box plot of attributes

Fig. 2 presents a comprehensive overview of a dataset through its five-number summary, showcasing the mean, median, and range of outliers for all parameters. The summary encompasses the minimum, first quartile, median, third quartile, and maximum values, offering insights into the presence of outlier data points. Moreover, it provides insights into the data's symmetry, clustering tendencies, and skewness degree [37].

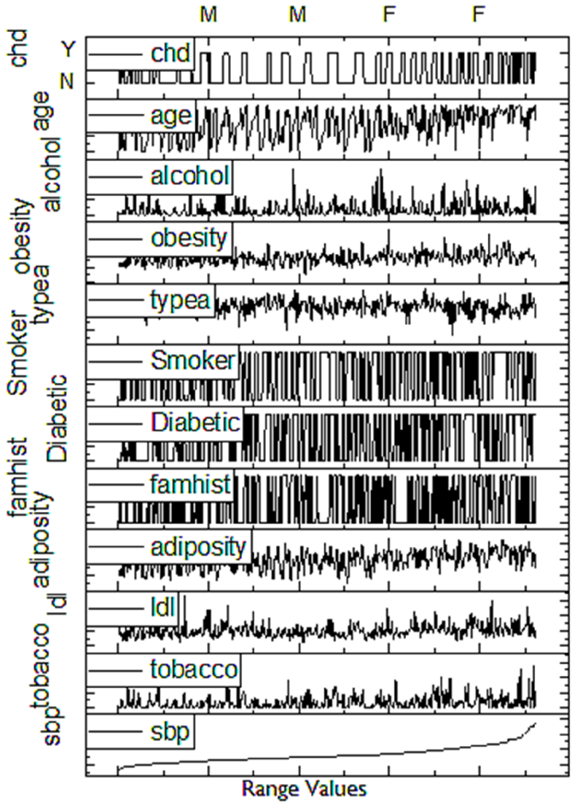

Fig. 3 Stacked bar graphs for cardiovascular data

Stacked graphs are particularly useful when the cumulative total is as important as the individual values. These graphs excel at presenting multiple values for different categories on bars or showcasing various values over time on lines. As a result, stacked graphs are commonly employed to effectively represent complex data patterns and relationships [38]. Thus, positive numbers are required for stacked graphs to function. Stacking bar graphs are used to illustrate how a bigger category is broken down into smaller categories and the impact that each component has on the overall sum as shown in Fig. 3.

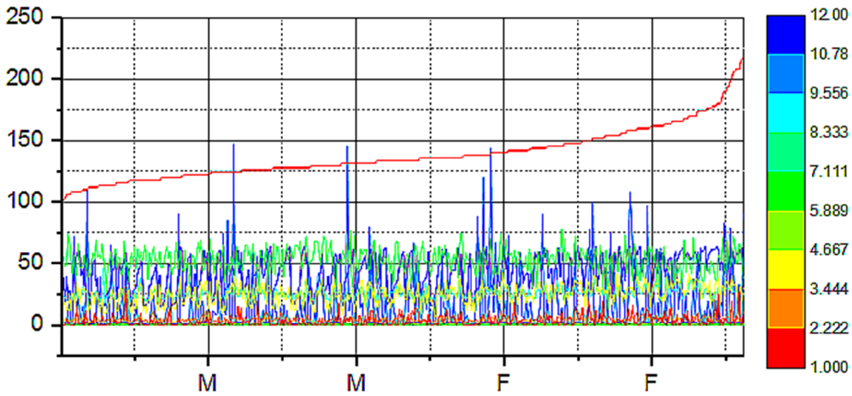

Fig. 4 Stacked line series graph for cardiovascular data

Fig. 4 depicts the overall line series representation of all of the attributes used in this study with respect to the general attribute. It determines numerical data over a predetermined time period. Plots have a variety of visual dimensions that humans can efficiently and pre-attentively differentiate [39], which is their main benefit when used in data visualization. In addition, regardless of the type of textures utilized, the results are typically more captivating and aesthetic that are more beautiful and beneficial.

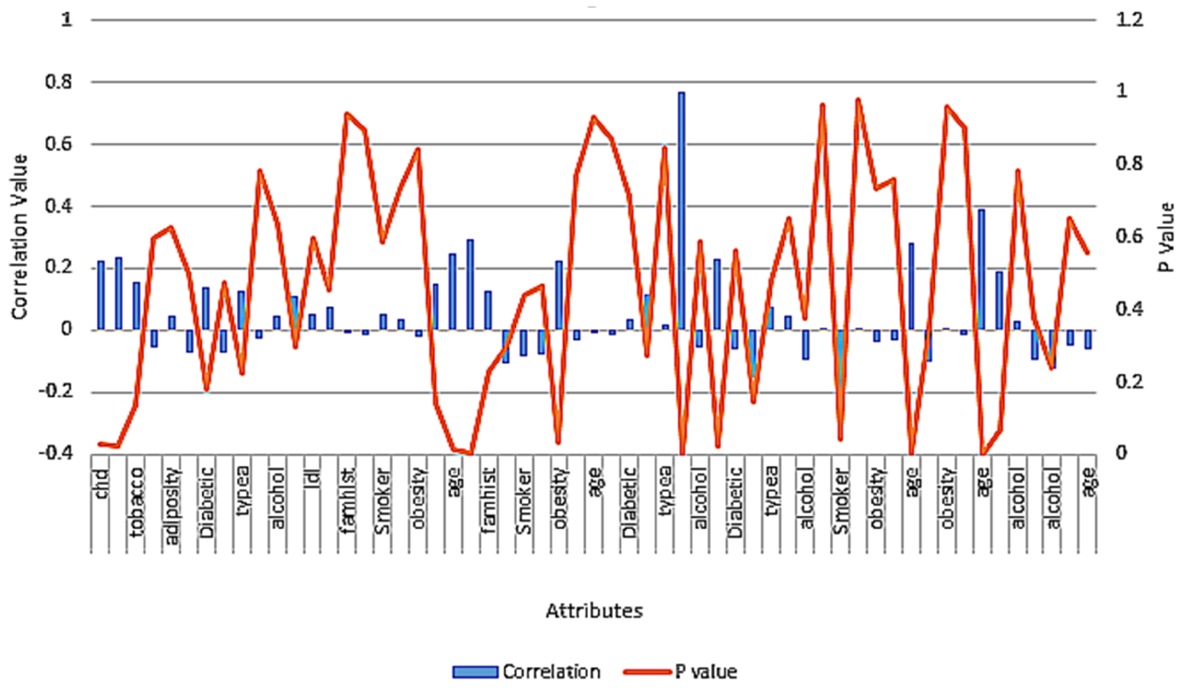

Fig. 5 Correlation and p value relation of attributes

The two most popular statistical tests for determining a link between variables, correlation, and p-value, are shown in Fig. 5. Correlation is used to determine whether there is any link between two variables, whereas p-value indicates whether an experiment's findings are statistically significant [40].

C. Preventing Overfitting in Cardiovascular Disease Prediction with L2 Regularization on Tabular Data

Since the data used in this paper is small, the proposed neural network model for predicting cardiovascular disease suffered from overfitting. Overfitting materializes when a model becomes excessively intricate, conforming to the noise present in the training data rather than capturing the fundamental patterns. To mitigate this concern, we implemented L2 regularization on the model's weights. This regularization technique helps prevent overfitting by imposing a penalty on overly complex model parameters.

Regularization techniques such as L1 or L2 regularization are effective at preventing overfitting on tabular data, as they can help the model learn more generalizable patterns by discouraging it from relying too heavily on specific features. Dropout can also be effective in some cases, particularly when the model has many layers or many parameters. L2 regularization is added to the weights of the first two layers of the neural network, with a regularization strength of 0.01. The Sequential model is compiled with the "adam" optimizer and binary cross entropy loss, and early stopping is used to stop the training process when the validation loss stops improving for 10 epochs.

By adding L2 regularization to the neural network weights, the model is encouraged to learn more generalizable patterns from the data, which can help prevent overfitting. We have used this method to solve the problem of overfitting in our neural network model for predicting cardiovascular disease. Algorithm 1 shows L2 regularization to tackle out the overfitting problem.

Algorithm 1, load_cardiovascular_data is a function that loads the tabular dataset containing information about cardiovascular disease patients. train_test_split is a function from scikit-learn that splits the dataset into training and testing sets, with a 80:20 split ratio. The Sequential model is defined with the Dense layers, with the first two layers having L2 regularization applied to their weights. The model is then compiled with the adam optimizer, binary cross-entropy loss, and accuracy metric.

Finally, the model is trained using the fit method, with the training data, number of epochs, batch size, and validation data specified. Early stopping is used as a callback to stop the training process when the validation loss stops improving for 10 epochs.

Algorithm 1
L2 Regularization for overfitting problem

```
x, y ← load_cardiovascular_data()
x_train, x_test, y_train, y_test ← train_test_split(x,y,test_size = 0.2, random_state = 42)

model = Sequential()
model.add(Dense(32, activation = ‘relu’, kernel_regularizer ← regularizers.l2(0.01),
input_dim(x_train.shape[1]))
model.add(Dense(16, activation = ‘relu’, kernel_regulaizer ← regularizers.l2(0.01))
model.add(Dense(1, activation = ‘sigmoid))
model.compile(optimizer= ‘adam’, loss=’binary_crossentropy’, metrics=[‘accuracy’])
history ← model.fit(x_train, y_train, epochs= 100, batch_size=32, validation_data= (x_test, y_test),
callbacks= [EarlyStopping(patience=10)])
```

D. CVD Prediction Model Using the Proposed GWLM-NARX Approach

The primary goal of the article is to create an adaptive CVD system utilizing the NARX neural network. The CVD prediction model receives the CVD data as input to forecast the cardiovascular disease. The NARX neural network is the adaptive model used to forecast cardiovascular diseases. It functions based on the error detected between the projected output and the actual data. The suggested GWLM technique is used to modify the weights of the NARX model based on the error. The Stanford CVD data are used for the adaptive prediction [41].

1. Autoregressive Models

AR models try to predict a series purely based on historical data or delay. Those AR models of order 1 that simply consider past values or delays are sometimes referred to as AR models. Mathematically, as Equation (1) illustrates as follows:

$$X_t = \gamma + \phi(X_{t-1}) + e_t \qquad (1)$$

where,
X_t = final output (target),
X_{t-1} = trailed output value (lagged target),
e_t = error,
γ & ϕ = the intercept and coefficient factors.

Recursive techniques are implemented using AR models, which are often referred to as long-term models since they start at the beginning of the data. These models are described mathematically as:

$$X_t = \gamma + \phi(X_{t-1}) + e_t \qquad (2)$$
$$X_{t-1} = \gamma + \phi(X_{t-2}) + e_{t-1} \qquad (3)$$
$$X_t = \gamma + \phi(\gamma + \phi(X_{t-2}) + e_{t-1}) + e_t \qquad (4)$$
$$X_t = \frac{\gamma}{1-\phi} + \phi^t X_t + \phi^{t-1}e_2 + \phi^{t-2}e_3 + \cdots + e_t \qquad (5)$$

where, $|\phi| < 1$.

The initial observation in this instance $(\phi^t X_t)$ is significant. In a roundabout way, the initial observation still has some influence. The entire purpose of stationery is to make this impression disappear [42].

2. NARX Model Architecture

Based on the availability of cardiovascular data, open model architecture was employed in this study's training phase. The use of this paradigm has two key benefits viz:

FFNN is more accurate when actual values are used as input, and multilayer perceptron (MLP) network training procedures can be utilized as usual. Additionally, this model is changed to a closed architecture after the training phase, which is better for multi-step-ahead prediction.

This section outlines the structure of the proposed model, which comprises three tiers: input, output, and hidden layers. Within the input layer, three types of vectors play a pivotal role in capturing diverse factors: the exogenous input vector, the delayed exogenous input vector, and the regressed output vector. These vectors collectively contribute to the model's ability to capture and interpret various influencing factors. Equation 6 below illustrates the mathematical problem formulation for the NARX model:

$$X(t) = f\big(z(t-1), \dots z(t-d), X(t-1), \dots . X(t-d)\big) \quad (6)$$

where $(t-d)$ is the cardiovascular data and $\mathrm{X(t)}$ is the result of the NARX model's disease prediction. By choosing the best weights for the input exogenous vector, our model performs significantly better and successfully predicts the CVD. Then, based on the error created between the anticipated output and the actual output, these weights are adjusted in accordance with the suggested GWLM method.

The results obtained from implementing the AutoRegressive (AR) and Nonlinear AutoRegressive with eXogenous inputs (NARX) models in predicting cardiovascular disease have a significant influence on the final outcomes. By utilizing past observations of a data with other input variables, the models are capable of making predictions about the likelihood of developing cardiovascular disease.The AR model uses a linear combination of past observations to predict future observations, and the NARX model is capable of capturing complex nonlinear relationships between the input variables and the output variable. In both cases, we utilized a training dataset to estimate the model parameters and validated the models using a separate validation dataset.

The results obtained from implementing the AR and NARX models allowed us to assess the risk of developing cardiovascular disease and develop personalized treatment plans for patients. The models identified the key risk factors and predicted the probability of developing cardiovascular disease based on a patient's current health status and lifestyle habits. Furthermore, the results allowed us to monitor the progression of the disease and evaluate the effectiveness of different treatments. Thus, implementing the AR and NARX models provides valuable insights into the potential risk factors and predictive factors contributing to cardiovascular disease, allowing for informed clinical decision-making and improved patient outcomes.

The model is evaluated by using the GWLM method to obtain optimal weights for the NARX model. This hybridization of GWO and LM algorithms enables adaptive prediction. The GWO algorithm's straightforwardness and minimal storage demands result in quicker implementation and processing speed in contrast to the Gauss-Newton (GN) algorithm. The GWLM-NARX NN model offers several advantages, including effective learning and faster convergence compared to other models. The study employs a comprehensive stepwise methodology that encompasses data collection, pre-processing, and cleaning stages, utilizing analytical tools such as principal component analysis (PCA). Subsequently, the implementation of the artificial neural network (ANN) approach follows, integrating testing, training, and validation procedures. Lastly, the cardiovascular data is subjected to the GWLM-NARX model, and the algorithm's effectiveness is assessed.

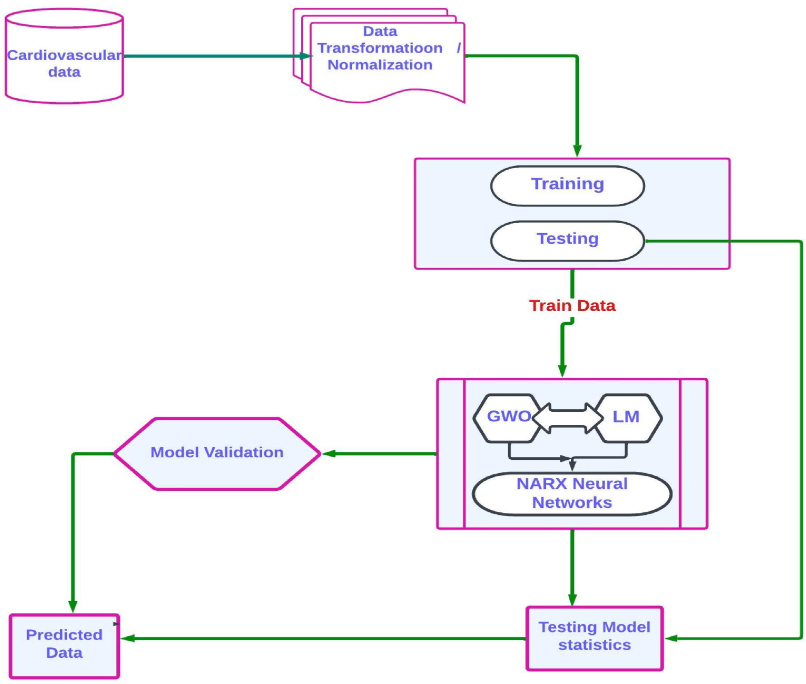

Fig. 6 Flow process of the model for cardiovascular disease prediction

Fig. 6 depicts the comprehensive step-by-step methodology utilized in this study. The sequence begins with data collection, followed by data pre-processing and cleaning stages, which incorporate diverse analytical techniques including principal component analysis (PCA). After implementing the ANN (Artificial Neural Network) methodology, which includes testing, training, and validation procedures, the GWLM-NARX model is applied to the data, and the algorithm's performance is subsequently evaluated. This evaluation allows for the determination of the accuracy and effectiveness of the model in predicting cardiovascular disease based on the input variables. Overall, the GWLM-NARX model provides a powerful tool for accurately predicting the likelihood of developing cardiovascular disease, leading to better-informed clinical decision-making and improved patient outcomes.

IV. RESULTS

In this part, the suggested paradigm for anticipating cardiovascular illness has been discussed. There is also discussion of a number of additional criteria, including experimental design and outcomes, database descriptions, and so on.

A. Database Description

A 1 M*N dimensional database including the original dataset's database is utilized to conduct the study of the cardiovascular dataset. These databases' specifics are provided in Table 4.

TABLE 4
DATASET TABLE STRUCTURE (DATABASE)

No.	Field	Type	Null	Default
1.	Gender	Char(2)	Yes	Null
2.	Systolic Blood Pressure (Sbp)	int(3)	Yes	Null
3.	Tobacco consume level (Cumulative)	double(2)	Yes	Null
4.	Ldl (low density lipoprotein cholesterol level)	double(2)	Yes	Null
5.	Adiposity	int(3)	Yes	Null
6.	Famhist (Family history disease of heart)	Boolean	Yes	Null
7.	Diabetic	Boolean	Yes	Null
8.	Smoker	Boolean	Yes	Null
9.	Typea (Type A)	int(3)	Yes	Null
10.	Obesity	int(3)	Yes	Null
11.	Alcohol (Consumption level)	double(2)	Yes	Null
12.	Age	Int(3)	Yes	Null
13.	CHD (coronary heart disease)	Boolean	Yes	Null

B. Validation Scheme

We used a holdout validation method to test the performance of our model. Specifically, we randomly divided the dataset into training and testing sets, where the training set was used to train the model and the testing set was used to evaluate its performance. Holdout validation is a useful and widely adopted method for evaluating the performance of machine learning models, and it can provide a good estimate of their generalization ability on the data. Holdout validation is a simple and popular validation scheme that involves randomly splitting the dataset into training and testing subsets and training the NARX model on the training subset and evaluating its performance on the testing subset. This validation scheme was found to be useful for estimating the generalization error of the NARX model.

The holdout validation method, as seen on Algorithm 2, was applied to test the effectiveness of a data-driven neural network for cardiovascular disease prediction that combines Grey Wolf Optimisation (GWO) and Nonlinear AutoRegressive with eXogenous inputs (NARX). Algorithm 2 assumes that the load_dataset function loads the input data X and corresponding output labels y from a file or database. The **train_test_split** function randomly splits the dataset into training and testing sets, with a test size of 0.2 and a fixed random state of 42 for reproducibility. The **num_splits** variable specifies the number of random splits to use for holdout validation, and the **mean_squared_error** function calculates the mean squared error of the predicted output values compared to the true output values. The NARX and GWO classes are assumed to be defined elsewhere in the code, with appropriate methods for defining the NARX model and optimizing its parameters using GWO.

C. Experimental Evaluation

This section presents a multitude of experimental findings pertaining to the likelihood of cardiovascular disease development. Fig. 7 illustrates the training performance of the neural network (NN), highlighting the optimal validation performance achieved at epoch 7, with a mean squared error (MSE) value of 0.1268. Fig. 8 displays the target & output response as well as the error between the anticipated and actual values. The findings are shown in Table 5 along with MSE and R values. The total regression values in each example range from 96.92% to 98.72%.

Algorithm 2
Holdout validation on cardiovascular data

```
x, y ← load_dataset()
x_train, x_test, y_train, y_test ← train_test_split(x, y, test_size=0.2, random_state=42)
narx_model = NARX()
gwo_algorithm ← GWO(narx_model.objective_function, narx_model.lb, narx_model.ub,
num_search_agents=50, max_iteration=100)
Mse_list = []
num_splits = 5

for i in range(num_splits):
        x_train, x_test, y_train, y_test ← train_test_split(x, y, test_size=0.2, random_state=i)
        best_solution, best_fitness ← gwo_algorithm.optimize(x_train, y_train)
        narx_model.set_parameters(best_solution)
end for

y_pred ← narx_model.predict(x_test)
mse ← mean_squared_error(y_test, y_pred)

mse_list.append(mse)
mean_mse ← np.mean(mse_list)
```

Fig. 8 displays the response graph between the training, validation, and testing goal and output with error. We can observe that the model fits the data the best since it consistently predicts the outcomes with the least amount of error. As a result, it provides the data's best model fit. Two-time delays and 15 hidden neurons were discovered to be the algorithm's most important parameters through trial and error.

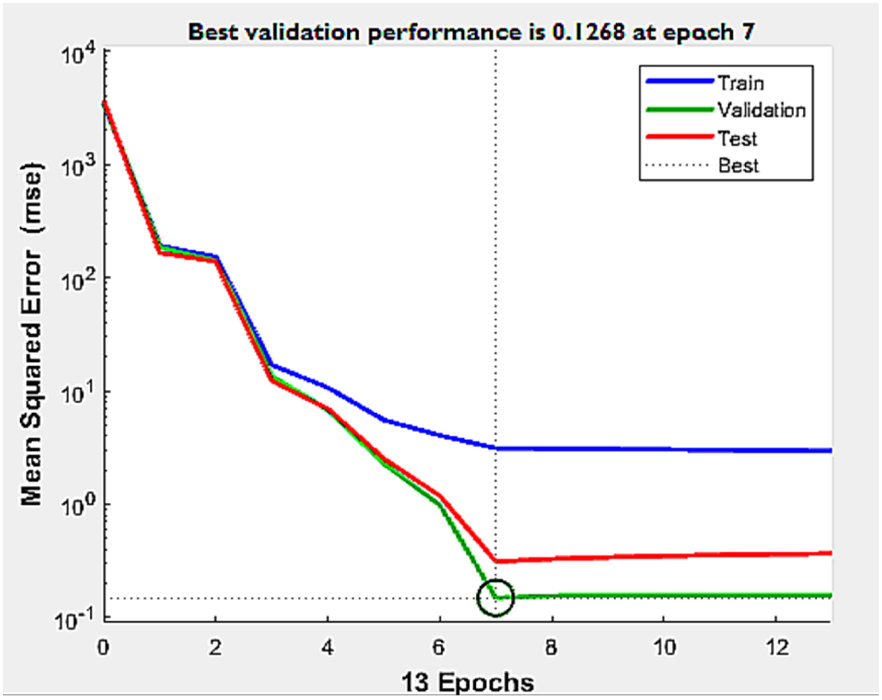

Fig. 7 Validation performance metrics

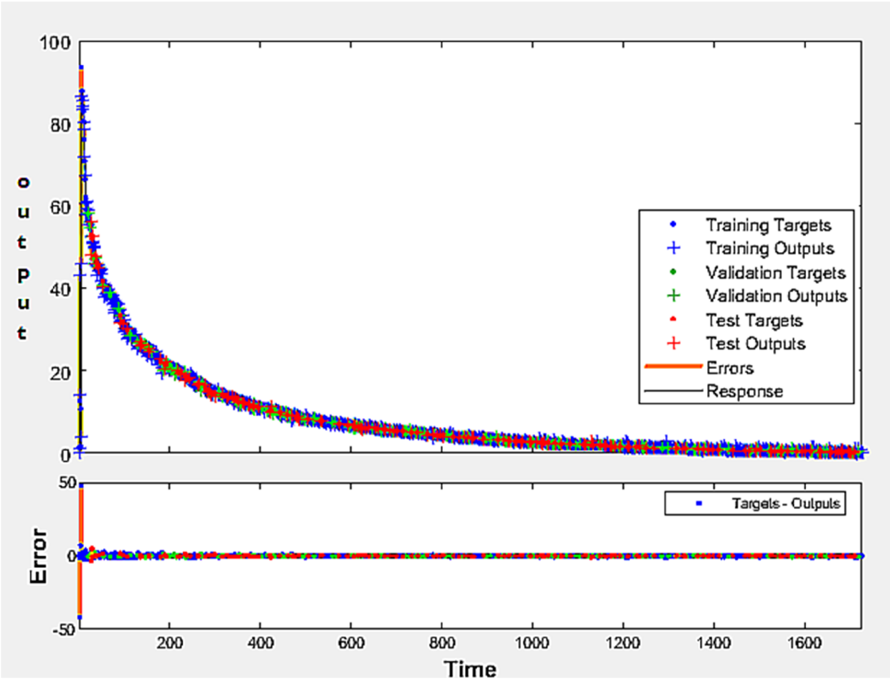

Fig. 8 Response of target outputs of training, validation and test outputs

D. Performance Metrics

The suggested model's MSE and R-values were examined in this study, and it was determined that the value of error needed to be reduced for successful outcomes. It was found that the training data's MSE value is 2.658%, with goal values of 982, and that the validation and testing data's MSE values are 0.137% and 0.28%, respectively. Additionally, all three examples' regression values stay between 96 and 99%. In the table below, the total accuracy statistics are displayed in Table 5.

TABLE 5
ACCURACY/ PERFORMANCE STATISTICS

	Target Values	MSE	R
Training	982	2.658	0.9692
Testing	152	0.137	0.9863
Validation	152	0.28	0.9872

The cardiovascular dataset has been subjected to the GWLM-NARX model, and as a result, prediction has greatly improved, and the mean square error rate has somewhat decreased. The execution performance of GWLM-NARX on MATLAB is marginally superior to that of conventional techniques. When GWLM-NARX is taken into consideration, this has a big effect on the results. However, how this is done depends on the particular dataset, and it may be different for other threshold datasets.

To calculate the accuracy of the model, we first need to convert the Mean Squared Error (MSE) values to Root Mean Squared Error (RMSE), as the RMSE is directly related to the accuracy metric. RMSE is calculated as the square root of MSE which can be seen on (7). Meanwhile the accuracy can be calculated as seen on (8).

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (7)$$

$$\text{Accuracy} = 100 - \text{RMSE} \quad (8)$$

Thus, the individual accuracy of each set can be calculated. First, training set is calculated where RMSE = √2.658 ≈ 1.63, so the accuracy is (100-1.63) ≈ 98.37%. After that testing set is also calculated which RMSE = √0.137 ≈ 0.37, then the accuracy is (100-0.37) ≈ 99.63%. Last, validation set is obtained by calculating RMSE = √0.28 ≈ 0.53 and the accuracy is (100-0.53) ≈ 99.47%. So, the accuracy of the model for the training set is approximately 98.37%, for the testing set is approximately 99.63%, and for the validation set is approximately 99.47%.

The overall model accuracy can be calculated by taking the average of the accuracies on the training, testing, and validation sets. Thus, the formula of overall model accuracy can be seen on (9). Therefore, the overall model accuracy is approximately 99.15%.

$$\text{Overall model accuracy} = \frac{(\text{Training Accuracy} + \text{Testing Accuracy} + \text{Validation Accuracy})}{3} \quad (9)$$

E. Comparative Analysis

This section of the study compares the GWLM NARX NN model with several conventional and ensemble approaches from the existing literature, including DT, RF, KNN, j48, CART as shown in Table 1. The MSE value for the GWLM NARX NN model ranges from 2.68% to 0.137%, while the overall performance of these conventional and ensemble machine learning techniques varies between 66% and 90%. These results demonstrate the effectiveness of our hybrid model, highlighting its promising performance compared to the classic and ensemble methods evaluated.

TABLE 6
COMPARATIVE ANALYSIS PERFORMANCE

Model	DT	LR	NB	KNN	HRFLM	RF	J48	AdaBoost	CART	GWLM_NARX	Overall Accuracy of GWLM-NARX Model
Accuracy	93.20	87.20	82.60	66.70	88.70	78.72	78.77	78.62	70.25	MSE 0.137 R 0.986	99.15%

In Table 6, we evaluated models using various evaluation criteria to give a thorough overview of model performance from various angles. For our proposed model, we employed mean squared error (MSE) and the R-value rather than accuracy, which was the main statistic for the majority of models. The precise objectives, purposes, and traits of our suggested model, as well as the requirement to take various performance factors into account, may provide justification for this decision. We wanted to provide a more detailed evaluation of model performance, take into account domain-specific factors, and give readers a benchmark by integrating several metrics.

V. DISCUSSION

The GWLM-NARX model emerges from the fusion of the Grey Wolf Levenberg model (GWLM) with the Nonlinear AutoRegressive with eXogenous Input (NARX) model. This hybrid model finds application in clinical cardiovascular data prediction and exhibits commendable performance across diverse scenarios. The dataset utilized in this study was divided into two distinct segments: the accurate sequence and the conclusive sequence. Specifically, the GWLM-NARX model was harnessed to align with the crucial sequence, encompassing the accurate outcomes within the dataset. To gauge its effectiveness, the GWLM-NARX model's performance was benchmarked against various conventional and ensemble techniques, such as decision trees (DT), classification and regression trees (CART), random forests (RF), and k-nearest neighbor (KNN).

The comparative analysis of our study with the [16] [13] [31] studies reveal variations in approaches and reported outcomes in predicting cardiovascular diseases using machine learning techniques. While all studies utilize similar parameters and work with the same dataset from Stanford University, they differ in the algorithms and techniques employed. In our research, we chose to use a total of 13 parameters as input for our prediction model. The decision to include these specific parameters was influenced by the fact that a previous study, referenced as [13] and [31], had also utilized the same set of parameters for their own prediction analysis. By adopting the same parameters as the previous study, we aimed to achieve consistency and comparability between the results. This approach allows us to build upon the findings of the earlier research, drawing meaningful comparisons and potentially validating or extending their conclusions. Moreover, using established parameters from prior work can provide a foundation of knowledge and help us understand how our new model performs in relation to existing approaches.

Study combines conventional learning algorithms and cutting-edge technologies, achieving noteworthy prediction accuracies with J48 and KERAS. Study [16] explores feature selection algorithms and showcases improved performance with both feature selection and deep learning techniques. Our study highlights the impressive performance of the Decision Tree algorithm and the GWLM-NARX model. However, it also emphasizes the challenges associated with feature selection and data complexity. Overall, the studies contribute to the understanding and advancement of cardiovascular data processing, but their approaches and reported results vary.

The results of the comparison revealed that the performance accuracy of all the techniques reached a plateau between 66% and 90%. However, the GWLM-NARX model exhibited enhanced accuracy statistics and reduced mean squared error (MSE) compared to the other methods. This improvement can be attributed to the GWLM-NARX model's adaptability and its ability to capture complex nonlinear relationships within the dataset. The implementation of the GWLM-NARX model resulted in increased accuracy statistics and a slightly lower MSE, indicating its effectiveness in predicting clinical cardiovascular data. Thus, based on the results presented that the GWLM-NARX model outperformed the conventional and ensemble methods, such as DT, CART, RF, and KNN, in terms of accuracy statistics and MSE reduction.

While the GWLM-NARX model showcased promising performance, it is essential to acknowledge the limitations and threats to the validity of the study. There are several restrictions on the study discussed that must be recognised. First off, the findings made strongly depend on the quality, representativeness, and size of the dataset that was employed. The results might not be applicable to real-world settings if the dataset is limited, not representative, or has biases. Furthermore, there is a chance of overfitting, where the GWLM-NARX model may have picked up on the quirks and noise unique to the training set, resulting in subpar performance on untrained data. In order to evaluate the GWLM-NARX model's dependability and generalizability, external validation of the study's results would be helpful. The lack of uncertainty analysis is also a problem because it prevents us from understanding how reliable and strong the model's predictions are. To offer a more thorough assessment of the GWLM-NARX model's performance, future research should concentrate on combining larger and more varied datasets, undertaking external validation, minimizing overfitting, and including uncertainty analysis.

VI. CONCLUSION

An intelligent algorithm can function as a useful tool and afterwards make a significant improvement to the accuracy of illness therapy. Through the utilization of traditional learning algorithms, ensemble methods, and state-of-the-art technologies, we gained a comprehensive comprehension of heart-related issues. This multifaceted approach led us to achieve notable prediction accuracy when working with cardiovascular data. Employing datasets sourced from the Stanford online healthcare repository, our initial steps encompassed the deployment of an array of classifiers, including decision trees (DT), J48, k-nearest neighbor (KNN), Naive Bayes, random tree, and random forest algorithms. With an amazing prediction accuracy of about 90%, DT outperformed other learning algorithms in terms of classification accuracy. The GWLM-NARX model, which outperformed both traditional learning classifiers and neural network technology in terms of prediction accuracy, came out on top. This success demonstrates the potency of our suggested prediction model, which made use of several machine learning methods and technologies. The peculiarities of the cardiovascular disease data were effectively displayed by these visualizations, which also supported the prediction mechanisms and offered insightful information. Our research also highlighted the difficulty in locating and choosing critical features for a precise diagnosis and the complexity of the data associated with cardiovascular illness. We were able to accomplish our study goals and significantly advance the field of cardiovascular data processing. Through the prediction of patient reactions to various medications, made possible by our study, personalised care was made possible in addition to assessing the likelihood of cardiovascular disease in a variety of patient populations. Moreover, by locating new biomarkers and diagnostic standards for the early recognition and diagnosis of cardiovascular illnesses, our research played a critical role in the development of novel diagnostic tools. Our research has expanded understanding of the fundamental causes of cardiovascular disorders by probing the complex dynamics of the circulatory system, potentially impacting the development of novel medicines and interventions.

As a result, our study reveals how the GWLM-NARX model was effective in helping us reach our research goals. The findings have important ramifications for patient care and management of cardiovascular disease. Given the intrinsic complexity of cardiac sickness data, future research efforts should focus on further improving accuracy and scalability in cardiovascular disease prediction systems. We have paved the road for a more precise and knowledgeable approach to cardiovascular data analysis and therapy through our thorough analysis and contributions. Our research ultimately aims to equip medical professionals with cutting-edge tools and insights, enabling them to make wise decisions and deliver better treatment to patients with cardiovascular disorders.

Author Contributions: *Sheikh Amir Fayaz*: Methodology, Programming, Writing -Original Draft, Writing -Review & Editing, *Majid Zaman*: Conceptualization, Methodology, Writing -Reviewing & Editing, Supervision. *Sameer Kaul*: Writing -Review & Editing. *Waseem Jeelani Bakshi*: Writing –Review & Editing.

All authors have read and agreed to the published version of the manuscript.

Funding: This research received no specific grant from any funding agency.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability: The data underlying this study are openly available at https://hastie.su.domains/ElemStatLearn/datasets/SAheart.data.

Informed Consent: There were no human subjects.

Animal Subjects: There were no animal subjects.

ORCID:
Sheikh Amir Fayaz: https://orcid.org/0000-0001-6606-0864
Majid Zaman: https://orcid.org/0000-0003-1070-8195
Sameer Kaul: https://orcid.org/0000-0003-0911-0073
Waseem Jeelani Bakshi: https://orcid.org/0000-0003-4848-0368

REFERENCES

[1] R.G. Nadakinamani, A. Reyana, S. Kautish, A.S. Vibith, Y. Gupta, S.F. Abdelwahab, and A.W. Mohamed, "Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques," *Comput. Intell. Neurosci.*, 2973324, pp. 1-13, 2022.

[2] I. Altaf, M.A. Butt, and M. Zaman, "Disease detection and prediction using the liver function test data: A review of machine learning algorithms," in *International Conference on Innovative Computing and Communications*, 2022, pp. 785-800.

[3] Rahim, Aqsa, Y. Rasheed, F. Azam, M.W. Anwar, M.A. Rahim, and A.W. Muzaffar, "An integrated machine learning framework for effective prediction of cardiovascular diseases," *IEEE Access*, vol. 9, pp. 106575-106588, 2021.

[4] A. Javeed, S.U. Khan, L. Ali, S. Ali, Y. Imrana, and A. Rahman, "Machine learning-based automated diagnostic systems developed for heart failure prediction using different types of data modalities: A systematic review and future directions," *Comput. Math. Methods Med.*, 9288452, pp. 1-30, 2022. https://doi.org/10.1155/2022/9288452

[5] I. Altaf, M.A. Butt, and M. Zaman, "Hard Voting Meta Classifier for Disease Diagnosis using Mean Decrease in Impurity for Tree Models," *Rev. Comput. Eng. Res.*, vol. 9, no. 2, pp. 71-82, 2022.

[6] A.S.M. Faizal, T.M. Thevarajah, S.M. Khor, and S.W. Chang, "A review of risk prediction models in cardiovascular disease: conventional approach vs. artificial intelligent approach," *Comput. Methods Programs Biomed.,* vol. 207, 106190.

[7] B. Martins, D. Ferreira, C. Neto, A. Abelha, and J. Machado, "Data mining for cardiovascular disease prediction," *J. Med. Syst.*, vol. 45, no. 1, pp. 1-8, 2021.

[8] K. Rajmohan, I. Paramasivam, and S. Sathyanarayan, "Prediction and Diagnosis of Cardio Vascular Disease-A Critical Survey," in *World Conference on Computing and Communication Technologies*, pp. 246-251, 2014.

[9] H.E. Bays, P.R. Taub, E. Epstein, E.D. Michos, R.A. Ferraro, A.L. Bailey, H.M. Kelli, "Ten things to know about ten cardiovascular disease risk factors," *American Journal of Preventive Cardiology*, vol. 5, 100149, 2021.

[10] B. Martins, D. Ferreira, C. Neto, A. Abelha, and J. Machado, "Data mining for cardiovascular disease prediction," *J. Med. Syst.*, vol. 45, no. 1, pp. 1-8, 2021.

[11] V. Dilsizian, H. Gewirtz, T.H. Marwick, R.Y. Kwong, P. Raggi, M.H. Al-Mallah, and C.A. Herzog, "Cardiac imaging for coronary heart disease risk stratification in chronic kidney disease," *JACC: Cardiovascular Imaging*, vol. 14, no. 3, pp. 669-682, 2021.

[12] S.A. Fayaz, M. Zaman, and M.A. Butt, "To ameliorate classification accuracy using ensemble distributed decision tree (DDT) vote approach: an empirical discourse of geographical data mining," *Procedia Comput. Sci.,* vol. 184, pp. 935-940, 2021.

[13] M. Ashraf, Y.K. Salal, S.M. Abdullaev, M. Zaman, and M.A. Bhut, "Introduction of feature selection and leading-edge technologies viz. tensorflow, pytorch, and keras: An empirical study to improve prediction accuracy of cardiovascular disease," in *International Conference on Innovative Computing and Communications*, pp. 19-31, 2022.

[14] S.A. Fayaz, M. Zaman, and M.A. Butt, "An application of logistic model tree (LMT) algorithm to ameliorate prediction accuracy of meteorological data," *Int. J. Adv. Technol. Eng. Explor.*, vol. 8, no. 84, pp. 1424-1440, 2021.

[15] S. Cristoforo, S. Rospleszcz, J.C. Kaiser, and K. Furukawa, "Heterogeneity in coronary heart disease risk," *Sci. Rep.*, vol. 12, no. 1 pp. 1-9, 2022.

[16] S.P. Rajamhoana, C.A. Devi, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, "Analysis of neural networks based heart disease prediction system," in *11th International Conference on Human System Interaction*, pp. 233-239, 2018.

[17] S. Babu, E.M. Vivek, K.P. Famina, K. Fida, P. Aswathi, M. Shanid, and M. Hena, "Heart disease diagnosis using data mining technique," in *International Conference of Electronics, Communication and Aerospace,* vol. 1, pp. 750-753, 2017.

[18] C. Krittanawong, H.U.H. Virk, S. Bangalore et al., "Machine learning prediction in cardiovascular diseases: a meta-analysis," *Sci. Rep.,* vol. 10, no. 1, pp. 16057-16111, 2020.

[19] G. Lippi, B.M. Henry, and F. Sanchis-Gomar, "Physical inactivity and cardiovascular disease at the time of coronavirus disease 2019 (COVID-19)," *Eur. J. Prev. Cardiol.*, vol. 27, no. 9, pp. 906-908, 2020.

[20] S. Aryal, A. Alimadadi, I. Manandhar, B. Joe, and X. Cheng, "Machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease," *Hypertension*, vol. 76, no. 5, pp. 1555-1562, 2020.

[21] I.K.A. Enriko, M. Suryanegara, and D. Gunawan, "Heart disease diagnosis system with K-nearest neighbors method using real clinical medical records," in *Proc. 4th Int. Conf. Frontiers Educ. Technol*., pp. 127-131, 2018.

[22] E. B. Randa, "An ensemble model for heart disease data sets: A generalized model," in *Proc. 10th Int. Conf. Inform. Syst.,* pp. 191196, 2016.

[23] S. Song, J. Warren, and P. Riddle, "Developing high risk clusters for chronic disease events with classi_cation association rule mining," in *Proc. 7th Australas. Workshop Health Inform. Knowl. Manage.*, vol. 153, pp. 69-78, 2014.

[24] A. Javeed, S.S. Rizvi, S. Zhou, R. Riaz, S.U. Khan, and S.J. Kwon, "Heart risk failure prediction using a novel feature selection method for feature refinement and neural network for classification," *Mob. Inf. Syst.,* 8843115, pp. 1-11, 2020. https://doi.org/10.1155/2020/8843115

[25] T. Sharma, S. Verma, and Kavita, "Prediction of heart disease using Cleveland dataset: A machine learning approach," *Int. J. Recent Res. Aspects,* vol. 4, no. 3, 1721, 2017.

[26] S.J. Al'Aref, K. Anchouche, G. Singh, "Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging," *Eur. Heart J.,* vol. 40, no. 24, pp. 1975-1986, 2019.

[27] A.H. Gonsalves, F. Thabtah, R.M.A. Mohammad, and G. Singh, "Prediction of coronary heart disease using machine learning: An experimental analysis," in *Proc. 3rd Int. Conf. Deep Learn. Technol.,* pp. 51-56, 2019.

[28] H.A.G. Elsayed and L. Syed, "An automatic early risk classification of hard coronary heart diseases using Framingham scoring model," in *Proc. 2nd Int. Conf. Internet Things, Data Cloud Comput.,* pp. 1-8, 2017.

[29] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542-81554, 2019.

[30] L. Yang, H. Wu, X. Jin, P. Zheng, S. Hu, X. Xu, W. Yu, and J. Yan, "Study of cardiovascular disease prediction model based on random forest in eastern China," *Sci. Rep.*, vol. 10, no. 1, pp. 1-8, 2020.

[31] M. Ashraf, S.M. Ahmad, N.A. Ganai, R.A. Shah, M. Zaman, S.A. Khan, and A.A. Shah, "Prediction of cardiovascular disease through cutting-edge deep learning technologies: an empirical study based on tensorflow, pytorch and keras," in *International Conference on Innovative Computing and Communications*, pp. 239-255, 2021.

[32] C. Perrino, P. Ferdinandy, H.E. Bøtker, B.J. Brundel, P. Collins, S.M. Davidson, and K. Ytrehus, "Improving translational research in sex-specific effects of comorbidities and risk factors in ischaemic heart disease and cardioprotection: position paper and recommendations of the ESC Working Group on Cellular Biology of the Heart," *Cardiovasc. Res.,* vol. 117, no. 2, pp. 367-385, 2021.

[33] A. Voutilainen, C. Brester, M. Kolehmainen, and T.P. Tuomainen, "Effects of data preprocessing on results of the epidemiological analysis of coronary heart disease and behaviour-related risk factors," *Ann. Med*., vol. 53, no. 1, pp. 890-899, 2021.

[34] S.A. Fayaz, S. Kaul, M. Zaman, and M.A. Butt, "An adaptive gradient boosting model for the prediction of rainfall using ID3 as a base estimator," *Revue d'Intelligence Artificielle*, vol. 36, no. 2, pp. 241-250, 2022. https://doi.org/10.18280/ria.360208

[35] M. Kenny, and I. Schoen, "Violin SuperPlots: Visualizing replicate heterogeneity in large data sets," *Molecular Biology of the Cell*, vol. 32, no. 15, pp. 1333-1334, 2022.

[36] B,A. Burton, B. Datta, and J. Spreer, "Flip graphs of stacked and flag triangulations of the 2-sphere," *Electron. J. Comb.*, vol. 29, no. 2, PP. 1-30, 2022.

[37] S.L. Turner, A. Karahalios, A.B. Forbes, M. Taljaard, J.M. Grimshaw, E. Korevaar, A.C. Cheng, L. Bero, and J.E. McKenzie, "Creating effective interrupted time series graphs: review and recommendations," *Res. Synth. Methods*, vol. 12, no. 1, pp. 106-117, 2021.

[38] S. Pirzada, and A. Altaf, "Line graphs of unit graphs associated with the direct product of rings," *Korean J. Math.,* vol. 30, no. 1, pp. 53-60, 2022.

[39] R. Mohd, M.A. Butt, and M.Z. Baba, "SALM-NARX: Self Adaptive LM-based NARX model for the prediction of rainfall," in *2nd international conference on I-SMAC (IoT in social, mobile, analytics and cloud)*, pp. 580-585, 2018.

[40] A.N. Nowbar, M. Gitto, J.P. Howard, D.P. Francis, and R. Al-Lamee, "Mortality from ischemic heart disease: Analysis of data from the World Health Organization and coronary artery disease risk factors From NCD Risk Factor Collaboration," *Circulation: cardiovascular quality and outcomes*, vol. 12, no. 6, e005375, 2019.

[41] J.Z.K. Toh, X.H. Pan, P.W.L. Tay, C.H. Ng, J.N. Yong, J. Xiao, and J.H. Koh, "A meta-analysis on the global prevalence, risk factors and screening of coronary heart disease in nonalcoholic fatty liver disease," *Clinical Gastroenterology and Hepatology*, vol. 20, no. 11, pp. 2462-2473, 2021.

[42] S.A. Fayaz, M. Zaman, and M.A. Butt, "A hybrid adaptive grey wolf Levenberg-Marquardt (GWLM) and nonlinear autoregressive with exogenous input (NARX) neural network model for the prediction of rainfall," Int. J. Adv. Technol. Eng. Explor., vol. 9, no. 89, pp. 511-524, 2022.