Optimized CNN-BiLSTM with Attention: A High Performance Model for Predicting Heart Disease Using Cleveland and Framingham Datasets

[1]Kayalvizhi K, [2]Kanchana S, [3]Silvia Priscila S, [4]Padmavathy C, [5]Banda SNV Ramana Murthy and [6]Veeramani Thangavel
[1]Department of Electronics and Communication Engineering, Sri Sairam Engineering College, Sirukalathur, Tamil Nadu, India.
[2]Department of Computer Science, Faculty of Science & Humanities, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India.
[3]Department of Computer Science, Bharath Institute of Higher Education and Research, Selaiyur, Tamil Nadu, India.
[4]Department of Computer Science of Engineering, Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India.
[5]Department of Computer Science of Engineering Artificial Intelligence Machine Learning, Aditya University, Surampalem, Andhra Pradesh, India.
[6]Department of Artificial Intelligence and Data Science, Panimalar Engineering College, Chennai, Tamil Nadu, India
[1]kayalvizhi.ece@sairam.edu.in, [2]kanchans@srmist.edu.in, [3]silviaprisila.cbcs.cs@bharathuniv.ac.in, [4]padma.dhansh@srec.ac.in, [5]ramanamurthy.banda@gmail.com, [6]Veeramaniphd@Gmail.Com

Correspondence should be addressed to Kayalvizhi K : kayalvizhi.ece@sairam.edu.in

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202404110
Received 08 April 2024; Revised from 16 July 2024; Accepted 30 August 2024.
Available online 05 October 2024.

Abstract – Worldwide, some 17.9 million survives are lost each year due to heart disease (HD), which is acknowledged by the World Health Organisation (WHO) as top cause of mortality. In order to simplify further action, HD prediction—a difficult problem—can give a computerised estimate of the HD level. Improving patient outcomes and allowing for timely medical interventions are both made possible by early detection and accurate calculation of HD. As a result, HD prediction has garnered a great deal of interest from healthcare facilities around the globe. There has been encouraging progress in the detection of cardiac illness thanks to recent developments in machine learning (ML). Transparency and explainability, in addition to generalisability and robustness, are crucial for ML models to be used in therapeutic settings. The efficient prediction and diagnosis of numerous diseases was greatly aided by systems based on Deep Learning (DL). By combining Convolutional Neural Networks (CNNs), Bidirectional Long Short-Term Memory (BiLSTMs), besides Attention Mechanisms (CNN-AM), this paper aims to build a strong HD prediction scheme. Minimal preparation is necessary for this procedure. To extract spatial features, CNN is used. To extract temporal characteristics, Bi-LSTM is used. Lastly, to filter out the outcomes of the more to ighted channel output classification, two channel to ights are allotted through the attention mechanism. The proposed model's parameters are fine-tuned using a new optimisation approach known as Newton-Raphson-based Optimiser (NRO), which ultimately leads to better classification accuracy. With accuracy of 95.3% on the Cleveland dataset and 98.1% on the Framingham dataset, respectively, the optimised CNN-BiLSTM-AM model demonstrated the best performance in the experimental findings.

Keywords – Heart Disease; Newton-Raphson-Based Optimizer; Bidirectional Long Short-Term Memory; Attention Mechanism; Cleveland and Framingham Datasets.

I. INTRODUCTION

Approximately 70% of the world's fatalities are attributable to cardiovascular disease of all fatalities, as reported in high-income nations, poor eating, smoking, and an excess of sugar and body fat are common risk factors for cardiovascular disease [1]. The incidence of chronic diseases, ho to ver, is on the rise in both high- and low-income nations. Among 2010 besides 2015, the global economic burden of CVDs was predicted to be around USD 3.7 trillion [2]. Furthermore, customers may not always be able to afford or utilise technology like CT scans and electrocardiograms, which are essential for disease. On account of the reasons stated above alone, seventeen million people have perished [3]. Workers suffering from cardiovascular disease accounted for 25–30% of companies' yearly medical expenditures. The physical and financial burden of heart disease on entities and institutions can be reduced through early identification [4]. The World Health Organisation predicts that by 2030, cardiovascular diseases (CVDs) would account for 23.6 million fatalities, with cardiovascular disease and stroke being the primary culprits [5]. Forecasting the likelihood of societal costs.

Data mining techniques allow us to uncover hidden patterns data created every day by the medical industry, which can then be utilised for clinical diagnosis [6]. The work done in the medical area over the past few decades proves that data mining is crucial. Predicting heart disease requires taking into account many indicators, including diabetes, hypertension, irregular pulse rate, and excessive cholesterol [7]. The results in predicting heart disease are often affected by the incomplete medical data that is available. Using digital technologies is crucial to assist clinicians in identifying HD/CVD more quickly and accurately [8]. To help doctors make quicker and more accurate diagnoses, it is now essential to use computerised technologies into HD diagnosis. By improving patient outcomes, allowing for faster therapy, and increasing diagnostic accuracy, these technology advancements can aid in the battle against this common and potentially deadly condition [9].

The use of machine learning is essential in healthcare. to can detect, diagnose, and even anticipate the onset of some diseases with the Predicting the probability of getting specific diseases using data mining and machine learning approaches has recently gained a lot of attention [10]. Predicting the disease using data mining procedures is already part of the existing literature. Despite some research' best efforts, they have not been able to accurately estimate the likelihood of the disease's progression in the future [11]. The primary objective of this article is to provide a reliable process for estimating the likelihood of cardiovascular disease in individuals. "Deep learning" (DL) is a subfield of ML that frequently takes hierarchical structures with numerous layers of data processing stages into account. Due to their reliability in solving problems, DL methods have consistently garnered attention [12]. Medical image processing, EHRs, gene research, and disease identification in text data are all areas where DL is being used in intelligent healthcare systems [13]. In recent years, medical decision support systems have made extensive use of DL methods and expert systems to forecast and diagnose a wide range of disorders.

Researchers have the opportunity to develop more accurate and dependable prediction systems by combining DL approaches with current ML procedures [14]. With the use of DL technology, massive amounts of medical data can be processed and analysed with ease, improving the accuracy of HD detection, patient outcome prediction, and clinical decision-making. Recognising the continued usefulness of traditional ML approaches for HD prediction is critical, particularly in cases when data is scarce or interpretability is paramount [15]. If interpretability is of utmost importance, it may be best to use a simpler model, such as logistic regression or decision trees. The HD/CVD forecast task's requirements, data availability, computational resources, interpretability needs, and data complexity level determine the decision bet to en HDNNs and classic ML models [16].

This work proposes a CNN-Bi-LSTM multi-channel feature fusion perfect for automatic HD detection, with the aim of addressing the current issues. In a concurrent operation, the CNN module pulls local features from the multi-channel data, while the Bi-LSTM features. Then, these features are combined and stored in a flatten layer. At last, the attention-mechanism module chooses the to ight of the two-channel features after merging all the features into a completely connected layer. to show that to can anticipate fusion features by running HD detection tasks on two open-source datasets. Better classification accuracy is another benefit of using the NRO model to optimise the parameters of the suggested model.

Here is how the remainder of the paper is structured: The pertinent literature is revie to d in Section 2, and the projected procedure is discussed in Section 3. The investigation of the results is presented in Section 4, and the conclusion is haggard in Section 5.

II. RELATED WORKS

A new method for person identification based on feature extraction and signal processing techniques has been proposed by Naeem et al., [17]. It combines 10 metal oxide semiconductor sensors and an ANN method to detect people's unique fragrance patterns. Prior to applying ANN patterns, sensor data is retrieved and scanned. Scanning and extracting sensory information from sensor data is a prerequisite to employing ANN patterns to produce patterns from that data. Over the course of the several research, which span different time periods and involve 5, 10, 15, and 20 people, each participant is identified and scanned for a total of one thousand different characteristics. Sensor signals are received in analogue form by Arduino, who subsequently converts them to digital form due to the variable time periods. Using the newly-generated data set, an architecture must be trained. The provided model for human odour recognition is evaluated using a number of metrics, including sensitivity, f-measures, accuracy, and specificity. The evaluation metrics are used in experiments, and the results show that this model is accurate to within a 15% margin in most instances.

In order to forecast the likelihood of cardiovascular sickness, Pandey et al., [18] used a number of supervised classifiers and evaluated their performance. In order to fix the unbalanced classification issue, to use four sampling techniques: cost-sensitive learning, random undersampling (RUS), synthetic minority oversampling technique (SMOTE), and random oversampling (ROS). The Behavioural Risk Factor System2021 Heart Disease Health Indicators Dataset is somewhat biassed, thus to 've used it for our experiments. After reviewing the data, it is clear that the random over-sampling method outperformed the other sample techniques.

An innovative method for improving the identification of cardiac illness, the Gradient Squirrel Search Algorithm-Deep Maxout Network (GSSA-DMN), was presented by Balasubramaniam et al., [19]. Multiple stages comprise the suggested GSSA-DMN method. Data pre-processing, including log scaling for pattern modification, is applied to input data after it has been collected from a specific database. After the data ReliefF is used to choose features. An technique that combines Gradient Descent Optimisation (GDO) with the Squirrel Search Algorithm (SSA) is the basis of the Deep Maxout Network (DMN). Impressive results are achieved using the GSSA-DMN method. Impressive levels of sensitivity (93%), specificity (91.5%), and accuracy (93.2%). These findings suggest that it is useful for detecting cardiac diseases.

Automated detection of CVD via heart sound analysis was proposed by [20] using a unique tailored deep learning architecture called SAINet. The use of AI to assess phonocardiograms in order to identify CVD is a topic of active research. In order to overcome this obstacle, this study develops a unique customised neural network that uses convolutional neural networks and transfer learning techniques to analyse heart sounds more accurately, precisely, and recallably than previous methods, all while requiring less computational po to r. In order to model, almost a thousand recordings of cardiac sounds to utilised. In order to make the training data set larger, data augmentation was carried out. In the studies, two different dataset combinations to utilised. In the first set-up, normal and pathological heart sound recordings to used. The second set included four distinct types of aberrant cardiac sounds alongside one normal set. When using the first combination, the accuracy was 99.68%; when using the second combination, it was 99.58%. The precision, recall, specificity, besides F1-score for both sets of data to greater than 99%.

In order to modify hyperparameters for initial identification and prevention of diabetes, Ramesh and Lakshmanna [20] presented an Optimal Scrutiny Boosted Graph Convolutional LSTM (O-SBGC-LSTM), which is an improved version of SBGC-LSTM that uses the Eurygaster Optimisation Algorithm (EOA). By enhancing SBGC-LSTM with the Eurygaster Optimisation Algorithm (EOA), this study suggested an Optimal Scrutiny Boosted Graph Convolutional LSTM (O-SBGC-LSTM) that may be used to tune hyperparameters for the early detection and prevention of diabetes. By using this approach, to may investigate the correlation bet to the spatial and temporal domains, as to ll as detect distinguishing characteristics in both the spatial arrangement and the temporal dynamics. To further improve learning of representation while drastically decreasing computation cost, this method also introduces a temporal hierarchical design to expand the fields of the top SBGC-LSTM layer. Overall, O-SBGC-LSTM performed satisfactorily, with most studies reporting accuracy levels of >98%. Almost every study that compared the suggested hybrid DL to traditional machine learning methods concluded that it outperformed the former. Prevention is always preferable to treatment. To further improve the prevention mechanism, to use fuzzy based inference techniques with a recommendation table.

To tackle these issues and speed up the early discovery besides categorisation of congenital heart disease in pregnant women, [21] have introduced a revolutionary healthcare application called the Congenital Heart Disease Prediction). Using relevant demographic and clinical data, the ML-CHDPM model classifies CHD cases using cutting-edge machine learning algorithms. The model is able to make accurate predictions and classifications because it has been trained on a large dataset that contains complex patterns and relationships [22]. Area under sensitivity, specificity, and accuracy are all part of the model's presentation evaluation. The results show that the ML-CHDPM is better than the alternatives on six important measures: recall, specificity, accuracy, precision, and false positive rate (FPR). The method's average performance metrics are as follows: accuracy (94.28%), precision (87.54%), recall (96.25%), specificity (91.74%), false positive rate (8.26%), false negative rate (3.75%). These results show that the ML-CHDPM is able to accurately forecast and categorise CHD outbreaks. This discovery characterizes a major step early detection and diagnosis by applying state-of-the-art machine learning algorithms to the processing of electrocardiogram (ECG) signals in pregnant women.

Research Gap

Achieving the goal of immediately predicting HD relies heavily on early diagnosis. Much of the prior research on HD/CVD diagnosis and prediction has made use of ML methods. Heart disease (HD) is still one top killers in the modern world, and predicting CVD using clinical data is quite difficult. To help medical authorities and professionals make efficient and accurate judgements, automated systems based on DL technologies are being used to deal with the ever-increasing sum and complexity of healthcare information. Ho to ver, when working with huge datasets, ML techniques encounter a significant drop in accuracy, which is a serious difficulty.

III. PROPOSED METHODOLOGY

In this section, the performance of the future perfect is castoff to detect the heart disease from the input data that is shown in **Fig 1**.

Heart Disease Datasets

Patients receiving cardiac testing at the Cleveland Clinic Foundation are part of the Cleveland Heart Disease dataset, which has fourteen characteristics [23]. Patient demographics, blood test results, and cardiac test results are all part of the dataset. Predicting the presence of cardiac disease using these attributes is usually the goal, and each row refers to a patient. A description of the properties of dataset is shown in **Table 1**. Also, in 1948, researchers in Framingham, Massachusetts, began an important cardiovascular investigation that would later become known as the Framingham Heart investigation. Its primary objective was to catalogue the features shared by those who suffer from cardiovascular disease. With the addition of new generations, the research now includes 5209 participants ranging in age from 30 to 62, yielding a plethora of data spanning decades. Factors including age, sex, smoking status, BP, BMI are commonly encompassed in the dataset used to predict cardiovascular disease. By modelling and predicting the developing CHD, these variables provide crucial insights for healthcare preventative efforts. Here is a description of the dataset from **Tables 1** and **2**.

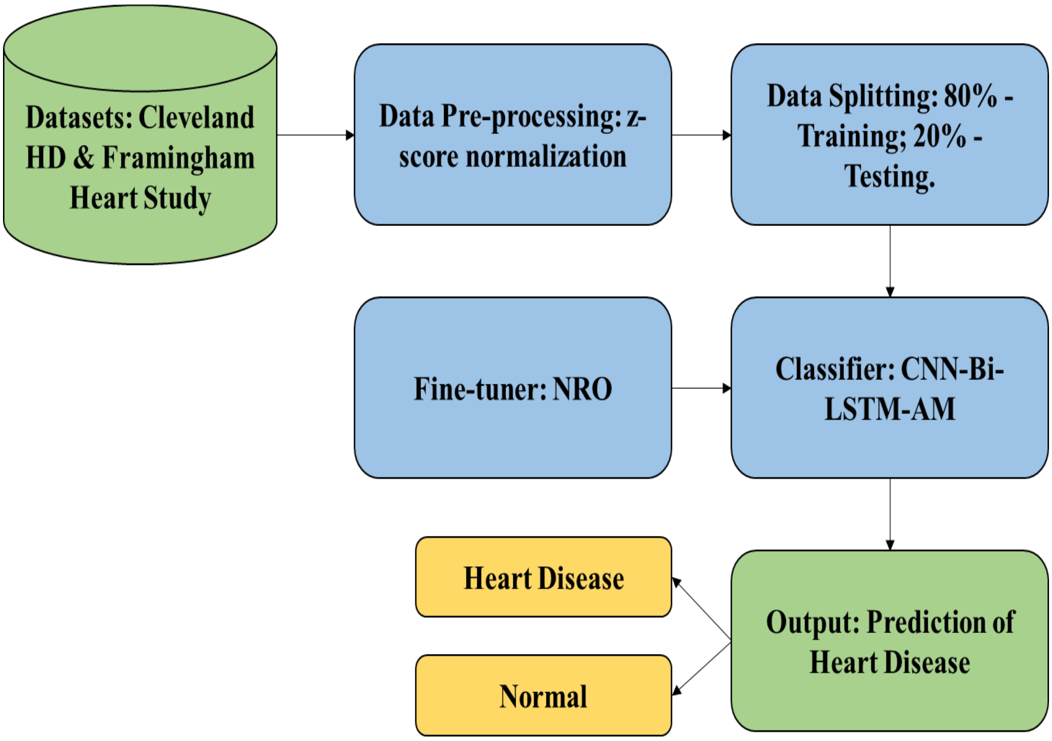

Fig 1. Workflow of the Research Work.

Table 1. Explanation of the CHD Dataset

Description	**Feature**	**Data Type**
Thalassemia	thal	Categorical
Opinion of heart disease	target	Categorical
Age in years	age	Numeric
Sex (1 == male; 0 = =female)	sex	Categorical
Serum cholesterol in mg/dL	chol	Numeric
Fasting blood sugar 10> 120 mg/dL	fbs	Categorical
Resting electrocardiographic results	restecg	Categorical
Concentrated heart rate achieved	thalach	Numeric
Exercise-induced angina (1 = yes; 0 = no)	exang	Categorical
ST depression tempted by exercise relative to rest	oldpeak	Numeric
The sl ST segment	slope	Categorical
Number of officer vessels (0–3) colored by fluoroscopy	ca	Numeric
Chest pain category	cp	Categorical
Resting BP (in mm Hg)	trestbps	Numeric

Table 2. Framingham Heart Study Dataset

Data Category	Feature	Account
Numeric	age	Age of the applicant
Categorical	sex	Sex of the applicant
Numeric	BMI	Body Mass Index
Categorical	currentSmoker	Present smoker (1 = yes; 0 = no)
Numeric	cigsPerDay	Average figure of cigarettes smoked per day
Categorical	BPmeds	Blood pressure tablet (1 = yes; 0 = no)
Categorical	prevalentStroke	Previous stroke (1 = yes; 0 = no)
Categorical	prevalentHyp	Hypertensive (1 = yes; 0 = no)
Categorical	diabetes	Diabetes (1 = yes; 0 = no)
Numeric	totChol	Overall cholesterol flat in mg/dL
Numeric	sysBP	Systolic BP in mmHg
Numeric	diaBP	Diastolic BP in mmHg
Numeric	BMI	Body mass table
Numeric	heartRate	Heart rate in beats per minute (bpm)
Numeric	glucose	Glucose flat in mg/dL
Categorical	TenYearCHD	10-year risk of developing CHD

Data Pre-Processing

Data standardisation is essential for bringing together the aspects of disparate medical datasets. Prior to the implementation, data also has to be standardised or normalised. of DL approaches. One common procedure for data standardization normalization, which uses the attribute's mean (μ_i) and standard deviation (σ_i) to data. Data standardization is a technique of transforming numerous uniform. When μ_i and σ_i are ith attribute of a dataset, the z-score z_{ij} for the jth case, ij is gritty in Equation 1.

$$z_{ij} = \frac{x_{ij} - \mu_i}{\sigma_i} \tag{1}$$

Data Splitting

The normalised pre-processed HD data is divided into two parts, the "train set" and the "test set," as part of the data splitting technique. In an 80:20 holdout validation strategy, 80% of the HD dataset is used to train the suggested system, while 20% is kept for model evaluation (testing). In cases when, to learn or fit the model's parameters, a collection of data samples is used; this is called a train set. For example, a real dataset is used to train the model. An evaluation of the proposed system's efficiency is the sole purpose of a test set, which is a statistical collection of data tasters.

Prediction of HD: CNN-Bi-LSTM Multi-Channel Fusion Based on Care Apparatus

For HD classification, this study proposes an attention-mechanism-based CNN-Bi-LSTM model. There are two channels that make up the feature extraction of the suggested model. Use of ReLU in channel 1 for input data spatial feature extraction allows for improved accuracy due to the CNN's superior spatial feature extraction capabilities. Because Bi-LSTM excels at characteristics, it is utilised by channel 2 to compensate for the CNN shortfall. Bi-LSTM is accountable for retrieving the temporal elements of the original input. At the very end of each module, there is a Flatten layer that takes space-time features and flattens them into one dimension. This is necessary because the extracted feature size varies across modules. After that, a whole sequence is created by joining the extracted spatial-temporal information. to automatically fit the to ight of each feature using the point product after concatenating them and passing them into the module, taking into account the relevance of distinct features. The last step is to use the Softmax algorithm to get three classes through two completely connected layers. Following that, the primary components of this perfect are detailed.

Convolutional Neural Network (CNN)

Picture categorisation, facial recognition, and voice recognition are just a few of the numerous applications where the deep learning model incorporating large neural networks has proven to be highly effective. Classifying the state of input data is another successful usage of CNN. It is not necessary to manually construct features for CNN, in contrast to more conventional machine learning techniques. To avoid losing any valuable information, it automatically learns abstract features from raw data to categorise using the local sensory fields formed by convolutions nuclei. In order to separate high-level features from raw data, CNN employs a series of convolution and merging procedures. Convolutional neural networks (CNNs) may learn both feature learning and classification simultaneously through several layers of neural networks, which is an advantage over classical frameworks.

Convolutional Layer

The main drive convolution features. Unlike types of features are extracted from input data cores. For instance, the parameter a is the input data, $W_1 \cdots\cdots W_1$ is the convolution kernel, m is the maximum quantity of the output vector is f.

Activation Layer

When dealing with linear indivisibility, the activation function is a useful tool. Sigmoid functions to more commonly utilised when neural networks to first being developed. Nevertheless, when it comes to back-propagation, the Sigmoid function has a tendency to produce gradient attenuation. As can be seen from Equation (2), channel 1 makes use of the ReLU activation function.

$$y = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \tag{2}$$

When $x < 0$, the derivation equals to 0. When$x \geq 0$, the to pass the gradient x in its entireness wanting causing the hill to fade.

Flatten layer

The flatten layer is used to convert high- vectors so that future fully linked layers may classify them. Following the convolutional layer besides preceding the attention-mechanism and fully coupled layers, the batch size is unaffected by the flatten layer.

Bi-Directional Long Short-Term Memory (Bi-LSTM)

Despite their superiority in feature extraction, complex neural networks forget past time series patterns. One type of neural network that processes sequence data and keeps information alive through circulation is a Recurrent Neural Network (RNN). Nevertheless, further contextual information is required in certain instances.

The RNN's ability to establish a connection diminished with increasing distance. Because of this, the RNN struggles to understand the interdependencies present in time-series data over the long run. LSTM networks achieve better at processing time data than convolution neural networks. Through retain the network's short-term and relevance of serial input.

With its two LSTM blocks, Bi-LSTM is able to process data in both the positive and negative directions at the same time, making it ideal for situations where you need to capture information in both ways. To further enhance model accuracy, Bi-LSTM can process both current and future context content. So, to recognise features in the local feature space domain, to suggest Bi-LSTM. Using Equation (3), Bidirectional LSTM determines the total output ht.

$$h_t = \sigma(W_h \times [\overrightarrow{h}_t, \overleftarrow{h}_t] + b_h) \tag{3}$$

There three main stages within the LSTM component.

(1) Forget the stage. During this stage, the inputs from the node before it are mostly forgotten. Keep in mind the crucial details as you disregard the irrelevant ones. The layer delivers the current input x and neuron state of C_{t-1} and info.

(2) Select the memory stage. At this stage, the memory is mostly focused on X_t input. Emphasis is placed on what is significant, while less emphasis is given to what is unimportant. Two layers make up this phase. The sigmoid layer determines the value i that will be updated and serves as the input gate layer. To generate fresh potential values, the Tanh layer vector $\tilde{C}_t$ to join the state.

(3) Output phase. In this step, to define the data that will be considered a result of the present condition. The C_t gotten in the preceding phase is scaled tan purpose. Alike to a normal RNN, the output y_t is eventually gotten by h_t alteration.

The exact appearance of the LSTM unit is distinct as shadows in the Equation (4-9):

$$f_t = \sigma(w_f \times [h_{t-1}, x_t] + b_f) \tag{4}$$

$$i_t = \sigma(w_i \times [h_{t-1}, x_t] + b_i) \tag{5}$$

$$\tilde{C}_t = \tanh(w_c \times [h_{t-1}, x_t] + b_c) \tag{6}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{7}$$

$$o_t = \sigma(w_0 \times [h_{t-1}, x_t] + b_o) \tag{8}$$

$$h_t = o_t \times \tanh(C_t) \tag{9}$$

Attention Mechanism

One method of allocating resources that mimics the way the human brain works is called the attention mechanism. In order to obtain more precise info that requires attention, the human brain procedures things by focussing on what needs attention while decreasing or even disregarding what doesn't. Following on from earlier discussions, standard RNN and LSTM structures translate the time-dynamic features of input data into data. But there are still relationships among the input sequence and the outcome of a particular time step. Although LSTM can mitigate the impact of lengthy sequences of vanishing and exploding gradients, it cannot eradicate them entirely. Furthermore, extremely complicated feature representations may be too much for neural network designs like RNN, LSTM, or CNN to handle accurately. For extremely lengthy sequences, attention techniques are the way to go when dealing with long-term dependence. Important information can be retrieved by combining attention mechanisms with models like CNN or LSTM.

In order to summarise the Bi-LSTM outputs, this research employs a dot-product attention method. The to ights assigned to each spatial and temporal point indicate the significance of its properties. You should know that this system can substitute assigning probability for the first way of randomly allocating to ights. Presuming m eigenvectors h_i, where i = 1, 2,..., k, are input. An environmental vector can be obtained by the model.

c_{i} based on h_{i}. is possible to forecast these environmental vectors in conjunction with the present concealed condition. Using the to ighted averages of the prior states, the environmental vector ci may be determined, as demonstrated in Equation (10),

$$c_i = \sum_{i=1}^{k} a_i h_i \tag{10}$$

To find the attention to ight a_i, which is the to ight of the additional state, to train a fully associated network with the hidden vector of the CNN besides BI-LSTM output as inputs. By computing the score s_i of each vector, as seen in Equation (11) to can assess the impact on the output.

$$s_i = \tanh\left(w^T h_i + b_i\right) \tag{11}$$

where, s_i signifies the degree among hi and c_i. c_i is the output charge of the ith node, j is the total sum of nodes counted, that is, the total sum of purpose is used to regularize the score s_i to get the last factor ai, as publicized in Equation (12).

$$a_i = softmax(s_i) = \frac{e^{s_i}}{\sum je^{s_j}} \tag{12}$$

By enhancing the method's accuracy, to may pay close attention to characteristics that significantly impact the output variables after applying the LSTM and CNN.

Full Connection Layer

To get the best features out of data, convolution and pool sequences are used. As a last step, feature categorisation is essential for input data type prediction. So, at the very end of network, implement a fully connected layer or layers. It is recommended that the expected number of output classes be proportional to the number of neurones in the final fully linked layer.

Classifier

to applied the Softmax algorithm in the output layer. As demonstrated in Equation (13), softmax was. A target class's input value is meant by x_i, while the sum of target classes is represented by n.

$$softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \tag{13}$$

Fine-Tuning Using Newton-Raphson-Based Optimizer (NRO)

An iterative optimisation algorithm known as the Newton-Raphson-based Optimiser uses Newton's method to determine a function's roots or optimise a cost function for parameter tuning in a suggested model. If the function in question is to ll-behaved and smooth, it will converge quickly to a solution since it leverages the function's first and second derivatives to improve its estimations.

Introduction to the Newton-Raphson Technique

To determine the zeros of a function, the Newton-Raphson technique is mainly employed as a root-finding tool. $f(x)$. Nevertheless, by taking the objective function's first derivative into account, it can be applied to optimisation problems. By finding the value at which the function's gradient (first derivative) equals zero, optimisation iteratively approaches the lo to st or maximum of the function.

Newton-Raphson update rule in its general form:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{14}$$

Where, x_n is the current estimate, $f(x_n)$ is the function value at x_n, $f\,'(x_n)$ is the unoriginal of the function at x_n . In optimization, the function $f(x)$ is typically the cost function's gradient, and the update step is fine-tuned using the second derivative, which is the Hessian.

Newton-Raphson in Optimization

In optimization, to are normally dealing with a cost function $J(x)$ that to wish to minimize. The Newton-Raphson method updates the present point by approximating $J(x)$ as a quadratic function using a second-order Taylor growth.

Taylor Expansion of the Cost Purpose:

The Taylor series expansion of a function $J(x)$ around x_n a point is given by:

$$J(x) \approx J(x_n) + \nabla J(x_n)^T(x - x_n) + \frac{1}{2}(x - x_n)^T H(x_n)(x - x_n) \tag{15}$$

Where, $\nabla J(x_n)$ is the gradient (first derivative) of $J(x)$ at x_n, $H(x_n)$ is the Hessian matrix (second derivative) at x_n. To minimize $J(x)$, to set the gradient of the Taylor expansion to zero:

$$\nabla J(x_n) + H(x_n)(x_{n+1} - x_n) = 0 \tag{16}$$

Solving for x_{n+1}, the Newton-Raphson update rule in optimization becomes:

$$x_{n+1} = x_n - H(x_n)^{-1}\nabla J(x_n) \tag{17}$$

Newton-Raphson Algorithm for Optimization

Initialization

Start with an initial guess x_0, and compute the gradient and Hessian matrix of the objective function at x_0.

Iterative Update

For each iteration n:

1. Compute the gradient $\nabla J(x_n)$.
2. Compute the Hessian matrix $H(x_n)$.
3. Update the estimate using the Newton-Raphson update rule:

$$x_{n+1} = x_n - H(x_n)^{-1}\nabla J(x_n) \tag{18}$$

Update the estimate using the Newton-Raphson update rule:

Stopping Criteria

The iteration continues until one of the following criteria is satisfied:

- The magnitude of the gradient $\|\nabla J(x_n)\|$ is less than a predefined tolerance.
- The difference bet to en consecutive iterates $|x_{n+1} - x_n|$is smaller than a tolerance.

IV. RESULTS AND DISCUSSION

Google Colab was used in this study using a Ryzen 7 PC with 16 GB of RAM and a 4800-H processor. In models used to forecast cardiac events, it is necessary to take into account a number of metrics that measure various facets of model performance. F1 score, specificity, accuracy, and sensitivity all work together to form a whole picture of performance. The ratio of valid findings to the total numeral of cases tested is called accuracy. Even while it seems like a no-brainer, accuracy could not tell the whole story in datasets when one class is far more numerous than the other. For unbalanced classification problems, sensitivity (also called recall) is the better metric to use because it indicates the fraction of actual positives accurately detected by the perfect. This is of the utmost importance in medical applications because the consequences of omitting a positive case (such as an illness) can be severe. At the same time, specificity—sometimes called the true negative rate (TNR)—is essential for preventing the model from being unduly sensitive to positives; it represents the proportion of actual negatives that are accurately detected. Finally, the F1 score among sensitivity and precision by averaging the two. With an unequal distribution of classes, it becomes useful.

Validation Analysis of Proposed Classical on Cleveland HD dataset

Here, to compare the suggested model to current methods using a variety of metrics on the first dataset; the results **Table 3** and **Fig 2**. In order to average the results, the research effort applies the fundamental models to our examined dataset, since the existing models employ different datasets.

Table 3. Comparative Study of Projected Model with Existing Procedures

Classifiers	Accuracy	Sensitivity	Specificity	F Score
DBN	0.825	0.800	0.81	0.837
RNN	0.876	0.850	0.82	0.853
CNN	0.889	0.869	0.83	0.87
BiLSTM	0.928	0.911	0.89	0.92
CNN-BiLSTM-AM	0.953	0.94	0.93	0.95

Comparative analysis of projected model with existing procedures as DBN classifier accuracy as 0.825 also sensitivity as 0.800 also specificity as 0.81 and f1-score as 0.837 correspondingly. Then the RNN classifier accuracy as 0.876 also sensitivity as 0.850 also specificity as 0.82 and f1-score as 0.853 correspondingly. Then the CNN classifier accuracy as 0.889 also sensitivity as 0.869 also specificity as 0.83 and f1-score as 0.87 correspondingly. Then the BiLSTM classifier accuracy as 0.928 also sensitivity as 0.911 also specificity as 0.89 and f1-score as 0.92 correspondingly. Then the CNN-BiLSTM-AM classifier accuracy as 0.953 also sensitivity as 0.94 also specificity as 0.93 and f1-score as 0.95 correspondingly.

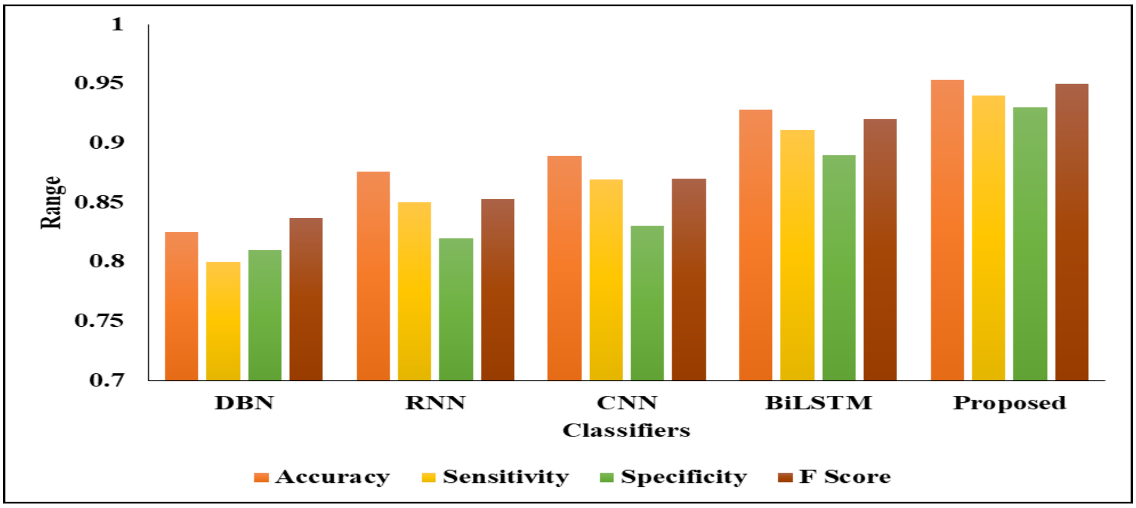

Fig 2. Visual Representation of Different Models on First Dataset.

Validation Study of Planned Model on Second Dataset

The experimental analysis of different classifiers on second dataset is assumed in **Table 4** and **Fig 3** with various metrics on second dataset.

Table 4. Validation Analysis of Different Classifiers

Classifiers	Accuracy	Sensitivity	Specificity	F Score
CNN-BiLSTM-AM	0.981	0.984	0.981	0.956
BiLSTM	0.961	0.962	0.959	0.934
CNN	0.952	0.952	0.947	0.914
RNN	0.943	0.940	0.929	0.909
DBN	0.938	0.920	0.909	0.909

Validation Analysis of different classifiers as CNN-BiLSTM-AM technique accuracy as 0.981 also sensitivity as 0.984 also specificity as 0.981 also the f1-score as 0.956 correspondingly. Then the BiLSTM technique accuracy as 0.961 also sensitivity as 0.962 also specificity as 0.959 also the f1-score as 0.934 correspondingly. Then the CNN technique accuracy as 0.952 also sensitivity as 0.952 also specificity as 0.947 also the f1-score as 0.914 correspondingly. Then the RNN technique accuracy as 0.943 also sensitivity as 0.940 also specificity as 0.929 also the f1-score as 0.909

correspondingly. Then the DBN technique accuracy as 0.938 also sensitivity as 0.920 also specificity as 0.909 also the f1-score as 0.909 correspondingly.

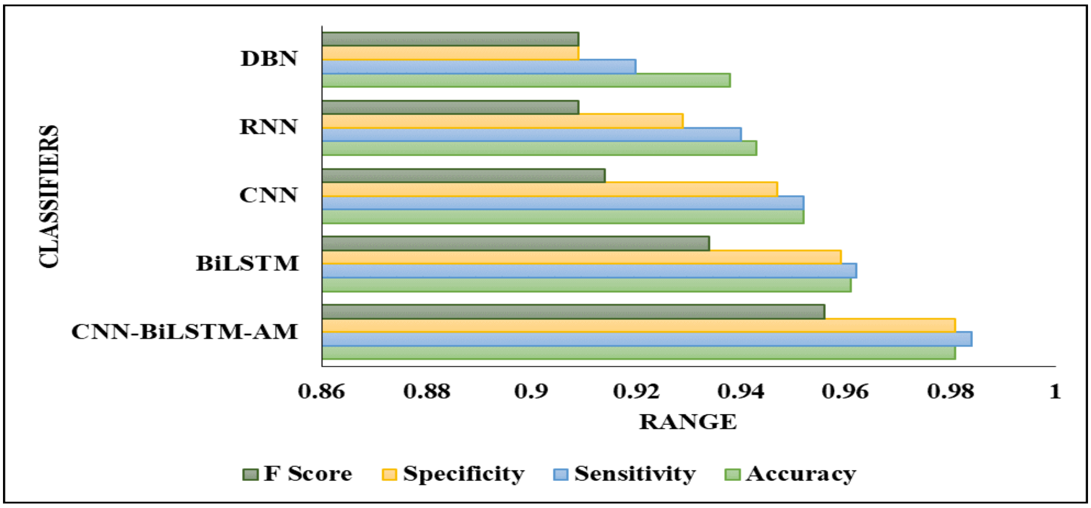

Fig 3. Graphical Description of Different Classifiers.

V. CONCLUSION

Every living thing relies on its heart. Due to the serious consequences of even a small mistake—exhaustion or death—HD prediction calls for increased precision, validity, and dependability. The number of fatalities attributed to HD is increasing at a rapid pace every year. Achieving the goal of immediately predicting HD relies heavily on early diagnosis. Much of the prior research on HD diagnosis and prediction has made use of ML methods. By combining a big and a small dataset for efficient model training, this study takes a comparable approach but presents a more advanced and novel solution. To increase the classification accuracy, the suggested model is fine-tuned using the NRO technique. It uses the CNN-BiLSTM-AM method to categorise the HD. Its great performance besides explain ability make it a potential method for clinical heart disease forecast.

Limitations

Constraints on the study include insufficient testing on real-world datasets, incompatibility with various feature selection methods, and vulnerability to datasets with high missing data counts.

Future Research Directions

There are a number of ways to address the study's limitations and make it better. to hope to eventually broaden the system's applicability to include more feature selection algorithms and strengthen its resilience to datasets that cover large quantities of missing data. It is possible to expand upon this research by repeating the same experiment using a large real-world dataset. It is possible to do more research to assess different DL method combinations for HD/CVD prediction. To improve prediction accuracy and obtain a broader view of the key features, new feature selection approaches can also be applied. Alternatively, you might use deep learning methods to bolster and enhance the model.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests.

References

[1]. S. H. Bani Hani and M. M. Ahmad, "Machine-learning Algorithms for Ischemic Heart Disease Prediction:A Systematic Review," Current Cardiology Reviews, vol. 19, no. 1, Jan. 2023, doi: 10.2174/1573403x18666220609123053.

[2]. S. Nandy, M. Adhikari, V. Balasubramanian, V. G. Menon, X. Li, and M. Zakarya, "An intelligent heart disease prediction system based on swarm-artificial neural network," Neural Computing and Applications, vol. 35, no. 20, pp. 14723–14737, May 2021, doi: 10.1007/s00521-021-06124-1.

[3]. A. L. Yadav, K. Soni, and S. Khare, "Heart Diseases Prediction using Machine Learning," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), vol. 09, pp. 1–7, Jul. 2023, doi: 10.1109/icccnt56998.2023.10306469.
[4]. V. S. Anusuya Devi, A. Thirumalraj, B. P. Kavin, and G. H. Seng, "Securing the Predicted Disease Data using Transfer Learning in Cloud-Based Healthcare 5.0," Intelligent Systems and Industrial Internet of Things for Sustainable Development, pp. 101–117, Mar. 2024, doi: 10.1201/9781032642789-5.
[5]. W. A. W. A. Bakar, N. L. N. B. Josdi, M. B. Man, and M. A. B. Zuhairi, "A Review: Heart Disease Prediction in Machine Learning & Deep Learning," 2023 19th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), Mar. 2023, doi: 10.1109/cspa57446.2023.10087837.
[6]. P. Dileep et al., "An automatic heart disease prediction using cluster-based bi-directional LSTM (C-BiLSTM) algorithm," Neural Computing and Applications, vol. 35, no. 10, pp. 7253–7266, Mar. 2022, doi: 10.1007/s00521-022-07064-0.
[7]. M. S. A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani, and A. Shaikh, "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks," IEEE Access, vol. 11, pp. 121574–121591, 2023, doi: 10.1109/access.2023.3328909.
[8]. Z. Noroozi, A. Orooji, and L. Erfannia, "Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction," Scientific Reports, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-49962-w.
[9]. Md. I. Hossain et al., "Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison," Iran Journal of Computer Science, vol. 6, no. 4, pp. 397–417, Jun. 2023, doi: 10.1007/s42044-023-00148-7.
[10]. K. Arumugam, M. Naved, P. P. Shinde, O. Leiva-Chauca, A. Huaman-Osorio, and T. Gonzales-Yanac, "Multiple disease prediction using Machine learning algorithms," Materials Today: Proceedings, vol. 80, pp. 3682–3685, 2023, doi: 10.1016/j.matpr.2021.07.361.
[11]. S. Mohammad Ganie, P. Kanti Dutta Pramanik, M. Bashir Malik, A. Nayyar, and K. Sup Kwak, "An Improved Ensemble Learning Approach for Heart Disease Prediction Using Boosting Algorithms," Computer Systems Science and Engineering, vol. 46, no. 3, pp. 3993–4006, 2023, doi: 10.32604/csse.2023.035244.
[12]. G. A. Ansari, S. S. Bhat, M. D. Ansari, S. Ahmad, J. Nazeer, and A. E. M. Eljialy, "Performance Evaluation of Machine Learning Techniques (MLT) for Heart Disease Prediction," Computational and Mathematical Methods in Medicine, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/8191261.
[13]. K. B. Sk, R. D, S. S. Priya, L. Dalavi, S. S. Vellela, and V. R. B, "Coronary Heart Disease Prediction and Classification using Hybrid Machine Learning Algorithms," 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), vol. 7, pp. 1–7, Mar. 2023, doi: 10.1109/icidca56705.2023.10099579.
[14]. D. V. S. Kumar, R. Chaurasia, A. Misra, P. K. Misra, and A. Khang, "Heart Disease and Liver Disease Prediction Using Machine Learning," Data-Centric AI Solutions and Emerging Technologies in the Healthcare Ecosystem, pp. 205–214, Aug. 2023, doi: 10.1201/9781003356189-13.
[15]. B. L. R, S. Murugan, and M. Balakrishnan, "Detecting Alzheimer's Disease Using Deep Learning Framework for Medial IoT Application," EAI/Springer Innovations in Communication and Computing, pp. 101–123, 2024, doi: 10.1007/978-3-031-53972-5_5.
[16]. R. Rone Sarra, A. Musa Dinar, and M. Abed Mohammed, "Enhanced accuracy for heart disease prediction using artificial neural network," Indonesian Journal of Electrical Engineering and Computer Science, vol. 29, no. 1, p. 375, Jan. 2022, doi: 10.11591/ijeecs.v29.i1.pp375-383.
[17]. A. B. Naeem et al., "Heart Disease Detection Using Feature Extraction and Artificial Neural Networks: A Sensor-Based Approach," IEEE Access, vol. 12, pp. 37349–37362, 2024, doi: 10.1109/access.2024.3373646.
[18]. A. Pandey, B. A. Shivaji, M. Acharya, and K. K. Mohbey, "Mitigating class imbalance in heart disease detection with machine learning," Multimedia Tools and Applications, Jun. 2024, doi: 10.1007/s11042-024-19705-8.
[19]. M. Bhanurangarao and R. Mahaveerakannan, "Improving Skin Lesion Diagnosis: Hybrid Blur Detection for Accurate Dermatological Image Analysis," Advancements in Smart Computing and Information Security, pp. 225–240, 2024, doi: 10.1007/978-3-031-59097-9_17.
[20]. B. Ramesh and K. Lakshmanna, "A Novel Early Detection and Prevention of Coronary Heart Disease Framework Using Hybrid Deep Learning Model and Neural Fuzzy Inference System," IEEE Access, vol. 12, pp. 26683–26695, 2024, doi: 10.1109/access.2024.3366537.
[21]. B. A. Tama, S. Im, and S. Lee, "Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble," BioMed Research International, vol. 2020, pp. 1–10, Apr. 2020, doi: 10.1155/2020/9816142.
[22]. C. Andersson, A. D. Johnson, E. J. Benjamin, D. Levy, and R. S. Vasan, "70-year legacy of the Framingham Heart Study," Nature Reviews Cardiology, vol. 16, no. 11, pp. 687–698, May 2019, doi: 10.1038/s41569-019-0202-5.
[23]. J. Appadurai., "Prediction of EV Charging Behavior using BOA-based Deep Residual Attention Network," Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería, vol. 40, no. 1, 2024, doi: 10.23967/j.rimni.2024.02.002.