Convolutional Neural Network and Bidirectional Long Short-Term Memory for Personalized Treatment Analysis Using Electronic Health Records

Prasanthi Yavanamandha[1, 3], D. S. Rao[2]
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad-500075, Telangana, India[1, 2]
Department of CSE-AIML & IoT-Faculty, VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad, India[3]

***Abstract*—Correct precision techniques have far not been introduced for modeling the modality risk in Intensive Care Unit (ICU) patients. Traditional mortality risk prediction techniques effectively extract the data in longitudinal Electronic Health Records (EHRs), that ignore the difficult relationship and interactions among variables and time dependency in longitudinal records. The proposed work, developed the Convolutional Neural Network – Bidirectional-Long Short-Term Memory (CNN-Bi-LSTM) method for personalized treatment analysis using EHR data. The CNN extracts the significant features from relevant features, focused on spatial-based relationships. Then, the Bi-LSTM layer captured the sequential dependencies and temporal relationships in patient histories that are essential to understand the treatment results. The Circle Levy flight – Ladybug Beetle Optimization (CL-LBO) integrates the circle chaotic map and Levy flight process in traditional LBO to select relevant features for classification. The proposed method reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% Area Under Curve (AUC) when compared to LSTM.**

***Keywords*—Bidirectional-long short-term memory; circle chaotic map; convolutional neural network; electronic medical records; Intensive Care Unit (ICU); ladybug beetle optimization**

I. INTRODUCTION

The healthcare centers are responsible for maintaining patient documents tracked to analyze the disease [1]. Manually handling the document is a difficult task, as the documents get lost, damaged, or destroyed [2]. Collecting confidential information during the readmission of patients became difficult and delayed the patient's treatment [3]. Securing the information of the medical data helps for the analysis of existing disease types, and maintains the confidential patient data [4-6]. Electronic Health Records (EHR) is a digital platform for the data storage and maintenance of medical patient history for future analysis [7]. The EHR allows the providers to access the details of the patients easily and quickly which provides more information of the patients to get better treatment [8-10]. The EHR platform also improves communication among healthcare that enhances coordination to reduce medical errors [11]. In the unavoidable and increasing digital transformation process of national healthcare system management, a significant size of structural EHR data is available [12].

The advancement in recent technologies has introduced many applications for the automatic recording and maintenance of medical patient history [13, 14]. Deep Learning (DL) and Machine Learning (ML) methods are introduced for automatic analysis of medical data. These methods undergo training to predict the result of data [15]. The Intensive Care Unit (ICU) technologies and automatic system development have effectively improved the efficiency of healthcare professionals including caregivers, nurses, and doctors [16]. The EHR is used for model training to learn details present in data [17]. The Medical Information Mart for Intensive Care III (MIMIC-III) dataset is considered as input for DL and ML methods. The study focused on solving the feature selection and dimensionality problems to maximize the performance of the model. The oversampling method is used in the pre-processing step to balance the data, fitness function-based optimization technique is used in the feature selection process to improve classification accuracy and prediction.

The essential contributions of the proposed framework are, the Synthetic Minority Over-Sampling Technique (SMOTE) is introduced in the pre-processing phase to balance the classes in data, which enhances the performance and generalization capability of the method, the Circle Levy flight – Ladybug Beetle Optimization (CL-LBO) based feature selection algorithm is developed to choose relevant and appropriate features whole feature subset, which helps to enhance the classification performance, the CL-LBO integrated the circle chaotic map and Levy flight in traditional LBO, which enhances the searchability and convergence rate of LBO to select relevant features, the Convolutional Neural Network – Bidirectional-Long Short-Term Memory (CNN-Bi-LSTM) is developed for classification to focus much on spatial-based relationships and to capture the sequential dependencies and temporal relationships in patient histories.

The previous research is analyzed in a literature survey is given in Section II. The proposed method explains the process of personalized treatment analysis using the proposed methodology presented in Section III. The results of the proposed model have been examined and compared with existing machine learning models in Section IV. Finally, the paper is concluded in Section V.

II. LITERATURE REVIEW

Fang Yang et al. [18] presented a deep morality risk prediction based on longitudinal EHR data. The LSTM model was used as a classifier to predict mortality rate. Visit-level and variable-level attention mechanisms were used to solve the gradient vanishing issue that occurred during the model training to predict appropriate outcomes that enhanced the efficiency of the model. However, the missing values presented in the data caused an overfitting problem during the process of training that affected the accuracy of prediction and reduced the performance of the model.

Shuai Niu et al. [19] presented a model to predict congestive heart failure mortality based on the feature selection method. Various ML techniques like Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), and Gradient Boosting (GB) were used as a classifier to predict the mortality rate. The Partial Swarm optimization (PSO)-based feature selection algorithms were utilized to select significant details of features that enhanced the accuracy of prediction and classification. However, the model struggled to interpret the pattern of features due to huge variables present in the data affected the training of the model to predict accurate outcomes and reduced the accuracy of classification. Maria Bampa et al. [20] presented a multimodal clustering technique for understanding the patient phenotype. A Multimodal Autoencoder (MMAE) model was used as a classifier to group the disease based on similar features. The model strength was enhanced by combining multiple modalities to improve the clustering of data and enhance the accuracy of classification. However, the model suffered from getting the approximate hyperplane required for the classification due to a huge number of variables that reduced the accuracy of classification.

Belel Alsinglawi et al. [21] presented a ML framework for the prediction of lung disease in hospitals. The RF algorithm was used as a classifier to predict the type of lung disease. The SMOTE was utilized to solve imbalance issues of the data that enhanced the training of the model to predict accurate outcomes. However, the model failed to identify the significant details of the data due to a dimensionality issue that reduced the accuracy of classification. Hua Shen [22] presented the Recurrent Neural Network (RNN) model to improve the prediction of diseases in healthcare. An attention mechanism was used to reduce the gradient vanishing problem that occurred during the training process where the maximum gradients were retained which increased the accuracy of prediction to get the approximate outcome. However, the model suffered from gradient vanishing issues where the gradients vanished during the training process which affected the accuracy of classification.

Sapiah Sakri et al. [23] presented a hybrid model to predict sepsis disease based on extracted features. Convolutional Neural Network (CNN) and Bi-directional LSTM models were combined to classify sepsis disease. The spatial and temporal features of the data were used to select the significant details of the data where the model identified the pattern of selected features that enhanced the performance of the models. However, the extra layers present in the neural network affected the interpretability of the model that reduced the accuracy of classification. Sarika R. Khope et al. [24] presented a featured engineering-based disease prediction Artificial Neural Network (ANN). The encoding method was utilized to eliminate the outer lines of the features and improve the accuracy of prediction and classification. The feature engineering technique was used to get significant details of the data that enhanced the accuracy of prediction. However, the overfitting issue occurred during the process because the irrelevant features present in the data fit the model and learned the noises that affected the accuracy of classification.

Chang Liu et al. [25] presented a different ML models like K Nearest Neighbor (KNN), DT, NB, LR, and RF for the early prediction of Multiple Organ Deficiency Syndrome (MODS). Kernel Shapley Additive exPlanations (SHAP) was utilized to enhance the quality of data. The ML methods used here enhanced the potential of the prediction by clustering the data which enhanced the accuracy of classification. However, the model struggled to identify the appropriate hyperplane required to classify the data approximately which reduced the accuracy of prediction and classification. Vinod Kumar Chauhan et al. [26] presented a DL-based attention model to enhance the classification of HER data. The LSTM model was used as a classifier along with a cross-attention-based transformer model for the prediction of the accurate disease. The LSTM model focused on identifying the significant features of data that enhanced classification accuracy. However, the model performance was reduced due to poor hyperparameter tuning due to a lack of efficient optimal value that reduced the accuracy of classification.

Ho-Joon Lee et al. [27] presented an ensemble consensus model based on HER for the classification of stokes. A stroke classifier technique was used to predict the outcome of the model. The Principal Component Analysis (PCA) technique was employed to minimize the dimensionalities of data and enhance the accuracy of classification. However, the model interpretability was challenging due to missing values presented in data that reduced the accuracy of classification.

From the overall analysis, the existing researches have limitations such as overfitting issues occurring during the process because the irrelevant features present in the data fit, dimensionality issues, struggle to interpret the pattern of features, and the missing values presented in the data caused an overfitting problem.

To mitigate these limitations, the CNN-Bi-LSTM method for personalized treatment analysis using EHR data is developed. The CNN extracts the significant features from relevant features, focused on spatial-based relationships. Then, the Bi-LSTM layer captured the sequential dependencies and temporal relationships in patient histories that are essential to understand the treatment results. The SMOTE technique is employed in the pre-processing phase to balance the classes in data.

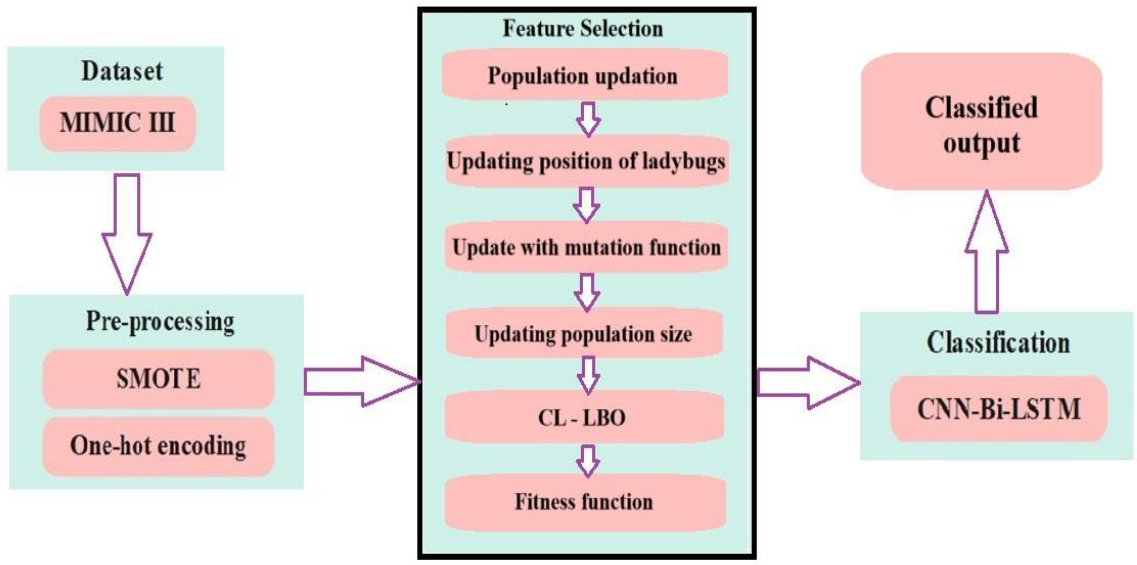

Fig. 1. Proposed architecture of personalized treatment analysis using EHR data.

The CL-LBO integrates the circle chaotic map and Levy flight process in traditional LBO to select relevant features for classification. This process improves the process of personalized treatment with high classification accuracy.

III. Proposed Method

The DL-based technique is developed in this work for personalized treatment analysis using EHR data. The MIMIC III dataset is used for personalized treatment and the data is pre-processed by using one-hot encoding and SMOTE techniques. Then, the relevant features are chosen by using CL-LBO that integrated the circle chaotic map and levy flight process. At last, the CNN-Bi-LSTM method is developed to classify features in data with classification accuracy. Fig. 1 describes the process of personalized treatment analysis using EHR data.

A. Dataset

The dataset used for this research is MIMIC III dataset [28] which includes 1177 cases for in-hospital mortality prediction and has a total of 51 features. The value of 0 represents life and 1 represents death. The BMI, gender, and Age are demographic factors studied. The essential signs are blood pressure, heart rate, respiratory rate, blood temperature, urine result, and saturation pulse oxygen. The characteristics of comorbidity include atrial fibrillation, depression, hyperlipidemia, hypertension, and diabetes. The white blood cells, basophils, lymphocytes, red blood cells, neutrophils, potassium, anion gap, sodium, bicarbonate, lactate, chloride, creatinine, calcium, magnesium, prothrombin time and creatine kinase are the laboratory variables.

B. Pre-processing

The pre-processing of data is employed to issue of imbalance samples is dissimilar from copy sample mechanism in random oversampling. The SMOTE technique synthesized the new samples among two minority samples by linear interpolation, which efficiently mitigates overfitting issue caused through random oversampling that makes balanced class distribution and enhances generalization capability of classifier [29].

The basic principle of SMOTE technique initially, choose every sample x_i from minority samples as the origin sample of a new synthetic sample, next, in accordance with up-sampling magnification n, randomly choose any neighboring samples k of similar classes in the sample x_i is an auxiliary sample in producing new samples, repeated n times, next linear interpolation is done among each sample and every auxiliary sample by using Eq. (1), and at last n synthesized samples are produced.

$$x_{new,attr} = x_{,attr} + rand(0,1) \times \left(x_{ij,attr} - x_{i,attr}\right) \quad (1)$$

In Eq. (1), the $x_{new,attr}$ for $attr = 1,2,\dots,d$ represents $attr - th$ attribute value in ith sample of the majority sample, the $rand(0,1)$ represents a random number between 0 and 1, the x_{ij} for $j = 1,2,\dots,k$ represents jth nearest neighbor sample of x_i, the x_{new} is the new sample synthesized among j and x_{ij}.

1) One-hot encoding: One-hot encoding is called as the one-bit effective encoding. This technique utilizes N-bit status registers for encoding N states. Every state has an independent register bit and one bit is valid. This technique is a representation of categorical variables in binary vectors with the benefit which transforming sample dataset to develop which is easy for utilizing the ML, particularly in classification algorithm. This process effectively enhances the measuring speed and method's performance.

C. Feature Selection

The general procedure of numerous metaheuristic algorithms is same to each other. In the proposed procedure, initial population of algorithms is sorted and evaluated depended on its evaluation. Next, population is updated and reevaluated.

After repeating the necessary updating and evaluation process for population, the optimal solution is described.

The LBO is motivated by coordination movements of ladybugs to identify the position with much heat. For this process, initial population that has $N(0)$ ladybugs are taken and the last population contains $N(k_{max})$ ladybugs and optimum fitness function is determined. Then, the LBO modeling is processed in three phases.

1) Population updation: The initial population includes $N(0)$ ladybugs that are positioned randomly in search area depended in uniform distribution. The ladybugs population is estimated with determined fitness function and sorted. Next, population moved to position with much heat in accordance with coordination movement. Because of the ladybug's nature which always moved in coordination with a swarm of ladybugs when searching for an appropriate position, the ladybugs follow each other by signals emitted through group members. Hence, much inclined to move towards it from ladybugs. In the proposed model, front ladybugs who have cable identify the position with more heat than others. For exploration and exploitation of the algorithm, the mutation phase is taken for certain individuals of the population that are randomly assigned to certain individuals in every iteration. However, at every phase of position update in population, its location in the search area is updated in accordance with other positions or mutation phases that are represented below.

2) Updating the position of other ladybugs: In every phase, entire positions of ladybugs are evaluated and updated. The old and new position of ladybugs are combined and optimal groups are selected in accordance with its fitness function values. The new population is assigned for updating and evaluating in following iteration. For updating every member of population in every iteration, another population member is chosen by utilizing the technique which is described. For instance, consider that goal is to update ith population of ladybug and jth population of ladybugs have selected for updating the member. Taken the location of ith ladybug in kth iteration is represented as $x(k)$. This ladybug moves in the outcome of three vectors for updating in $(k+1)$ iteration moves a little towards jth ladybug and move in the direction of jth ladybug towards $(j-1)th$ ladybug. To avoid local optima for movements of ith ladybugs and to maintain the balance between exploration and exploitation, the third direction is also taken, that is coefficient of their present position. All the determined movements are required to be multiplied through random values to sustain the random search. Moreover, the third movement that is assigned for avoiding the local optima is multiplied through the ratio of individual heat value to all heat values of the population. As an outcome, population members that are trapped in local optima have a chance to outflow local optima due to its high coefficient of third movement. The mathematical formula for the new position of ith ladybug is given as Eq. (2),

$$x_i(k+1) = x_i(k) + rand \times \left(x_j(k) - x_i(k)\right) + rand \times \left(x_j(k) - x_{j-1}(k)\right) + rand \times |C_i|^{-\frac{k}{N(k)}} \times x_i(k) \quad (2)$$

In equation (2), the C_i represents the same proportion of ith ladybird cost to total cost of entire ladybirds in kth iteration of developed algorithms. The mathematical formula for parameter value is measured by using Eq. (3),

$$C_i = \frac{f(x_i(k))}{\sum_{t=1}^{Nk} f(x_t(k))} \quad (3)$$

The Roulette-wheel selection is assigned for selecting the jth ladybug is utilized for updating the ith ladybug location using Eq. (3). In the general equation, for choosing jth ladybug from $N(k)$ ladybugs, the distance between 0 and 1 is separated to $N(k)$ unequal phases. Every phase corresponds to one of ladybugs and length of every phase is inversely relevance to fitness function of respective ladybug. For ladybug with much optimum fitness function, length of respective phase is higher. Next, the random number among 0 and 1 is selected. In present phase, random number determined in that phase of separation is positioned. The respective ladybug of the chosen phase is selected as jth ladybug. This is clear that ladybugs with warmer positions have a high chance of being selected. The P vector respective to the population with $N(k)$ ladybugs are determined. The Roulette-wheel selection expected a represents input vectors like P and its mathematical formula is given as Eq. (4),

$$P = \left[P_1, P_2, \dots, P_{N^k}\right], P_i = e^{-\beta \frac{f(x_i(k))}{f_{worst}}} \quad (4)$$

In Eq. (4), the β represents pressure coefficient in Roulette-wheel selection technique and f_{worst} represents worst value of fitness function for present iteration in the process of algorithm. The higher β, the optimal ladybugs of population have a good chance of being chosen in Roulette-wheel selection. This is clear that new location of ith ladybug is defined through outcome of three vectors r_1, r_2 and r_3.

3) Updation in accordance with mutation process: Considering the mutation is updation process of the population is critical for exploration of unidentified phases of search area and escapes from local optima. The mutation phase in search area causes the maximum speed of algorithm. Therefore, updation technique of every position of the ladybug includes in accordance with other ladybugs and mutation is randomly defined. Considers the ith ladybug is mutated. The amount of decision variables of ith ladybug is mutated and its mathematical formula is given as Eq. (5),

$$n_m = round(n \times \mu_m) \quad (5)$$

In Eq. (5), the μ_m represents mutation rate and the n represents length of decision variable. Hence, n_m represents variables presented in n variables of ith ladybug is chosen randomly. Next, random variables in the feasible phase are replaced for choosing location of ith ladybug.

4) Updating the number of population size: In searching for warm place that is common for ladybugs to disappear and lost.

The ladybug moves away from annihilate and others because of cold. The mathematical methods for the annihilation of ladybugs in search are taken in LBO. The mathematical formula for the number of ladybugs in different phases is measured by using Eq. (6),

$$N(k+1) = round\left(N(k) - rand \times N(k)\left(\frac{NFE}{NFE_{max}}\right)\right) \quad (6)$$

In Eq. (6), the NFE represents the number of function estimations and the NFE_{max} represents a maximum of NFE. Whether the amount of function evaluation is stopping criteria of LBO. Whether several iterations are a condition to terminate the algorithm. The mathematical formula for the ladybug in every iteration is given by using Eq. (7),

$$N(k+1) = round\left(N(k) - rand \times N(k)\left(\frac{k}{k_{max}}\right)\right) \quad (7)$$

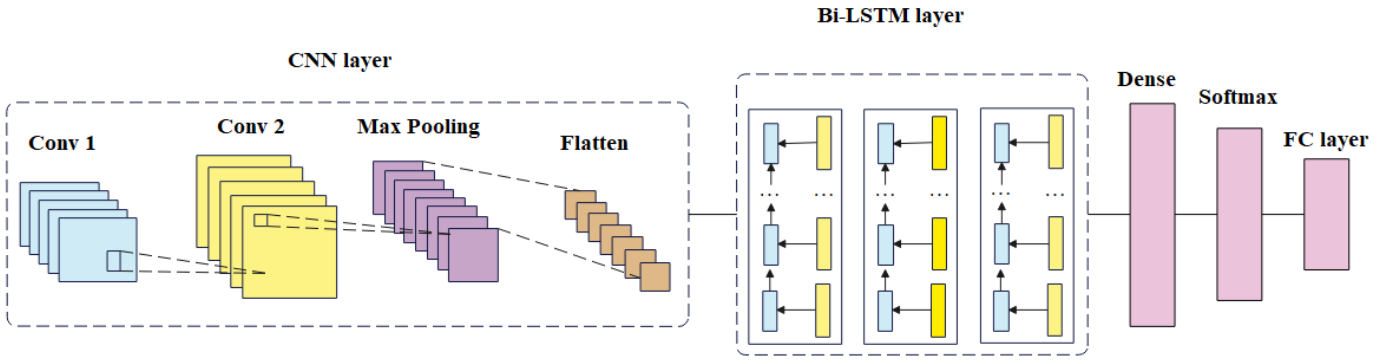

Fig. 2. Architecture of CNN-Bi-LSTM network.

In Eq. (7), the k represents iteration and the k_{max} represents maximum iteration.

5) CL-LBO: Enhanced non-linear LBO algorithm depended on Circle chaotic map and Levy flight process is introduced in the proposed work. Firstly, initialize the population by utilizing a circle chaotic map to maximize bee diversity. The integration of LBO and levy flight provides an algorithm with high capability in global exploration. Additionally, the non-linear adaptive weight operator is implemented for modifying the weight coefficient of bee following behavior in CL-LBO. The relationship between global exploration and local exploitation in the iteration process is effectively balanced. The chaotic values of actual circle operation are grouped in the range of [0.2, 0.5]. For making uniform chaotic value distribution, the mathematical method of the Circle chaotic mapping strategy is enhanced. The numerical expression for the circle chaotic map is given as Eq. (8),

$$x_{i+1} = mod(x_i + 0.2 - (0.5/2\pi)\sin(2\pi x_i), 1) \quad (8)$$

In Eq. (8), the x_i represents ith chaotic particle and the x_{i+1} represents $(i+1)th$ chaotic particle. The frequency histogram and plot of the initial candidate solution of the circle chaotic mapping process.

a) Levy flight: The trajectory and movement of several little insects and animals in life have Levy flight characteristics. The insects and animals include flies and ants. Numerous animals in natural usage of Levy flight strategy are the essential path of foraging. The Levy flight is the process compatible with Levy distribution. The step size of the Levy flight is mixed and random with short and long distances that make it easy to search the huge scale and with unknown scope compared with Brownian motion. In the searching procedure, the Levy process utilized little steps for walking and long steps for jumping, which allowed it for effective of local attraction points. Hence, in the random searching issue, numerous heuristic algorithms adopted this strategy for changing the iteration process that effectively supports the algorithm to get the influence of local attraction points. The mathematical formula for the strategy is given as Eq. (9),

$$L(s) \sim |s|^{-1-\beta} \quad (9)$$

In Eq. (9), the β in is range [0, 2], the s represents step size and the $L(s)$ represents the probability density of step size in accordance with Levy modeling.

6) Fitness function: For calculating the fitness value of generated CL-LBO agents, the Mean Square Error (MSE) fitness function is dependent on measuring the difference between original and predicted values through produced agents for training samples. The mathematical formula for MSE is in Eq. (10),

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y - \hat{y})^2 \quad (10)$$

In Eq. (10), the y describes the actual value, the $\hat{y}$ describes the predicted value and the n describes a number of instances in the training set. From the CL-LBO algorithm, the 21 selected relevant features are given to the classification phase for further process.

D. Classification

The classification is performed by using the CNN-BiLSTM network. Initially, the architecture is developed for leveraging the advantages of CNN to filter and remove the noisy data and obtain substantial data from time series. The noisy data is eliminated by using dimensionality reduction. Hence, inappropriate data (noise) is not involved in minimized matric. The significant data in the hidden interval is attained through employing the process of convolutional such as kernel matrix or filter passed by input matric, various characteristics based on

kernel and that obtained the hidden data. Next, the design is developed to pull the ability of the Bi-LSTM setup for modeling and forecasting both short and long-term dependencies in temporal data. In this design, the input data is processed twice, concurrently from left to right and from right to left. Both context readings are combined into results and provide much more comprehensive data of information context than unidirectional LSTM. Here, The CNN functions as the encoder that identifies and extracts features from the input data, whereas the BiLSTM assists as the decoder, analyzing historical dependencies in the data stream. The CNN block contains the CNN layer, pooling layer, and flattening process. The output from the CNN block is conceded as input to the BiLSTM block. This block contains a BiLSTM layer, a dropout layer, and a dense layer. The convolutional and pooling layers are included to mine features from the data, operating it in a matrix format. Fig. 2. demonstrates the design of CNN-BiLSTM network architecture.

1) CNN block: The convolution and pooling layer are used to filter the incoming information for extracting significant data from the matrix. The convolution process is done through a convolution layer among input and small matrices known as filters and kernels. Generally, numerous filters are included as sliding windows with a certain height and width. These filters slide across the input matrix with a specified stride, assigning a convolution operation to each overlapping sub-region of the matrix. This procedure produces a convoluted matrix that bags a specific feature of the original input matrix. By using multiple convoluted features, the technique delivers a more specific representation of the input matrix. Consider *H* as an input matrix, the *I* represents kernel matric and the indices m and n signifies the rows and columns of the subsequent matrix, respectively and its mathematical formula is given as Eq. (11) and Eq. (12),

$$R[m,n] = (H \cdot I)[m,n] \tag{11}$$

$$= \sum_j \sum_k I[j,k] \cdot H[m-j, n-k] \tag{12}$$

The convolution layer utilizes the ReLU activation function, which is a majorly used activation function. Its primary benefit is that it does not activate all neurons at the same time and transforms entire negative values to 0. Because of this, ReLU has high computation efficiency. The ReLU is six times quicker than other activation functions like tanh. The mathematical formula for the ReLU activation function is given as Eq. (13),

$$f(X) = x^+ = \max(0, x) \tag{13}$$

The max pooling layer will follow the convolution layer and turns as a sub-sampling technique to decrease the dimensions of the convolution matrix. It attains this by removing definite values while retaining key features recognized by each filter. For every patch of the matrix, it chooses the heights value, creating new matrices that helps as condensed notations of the convolution matrices. This pooling procedure improves the robustness of the technique. Finally, the pooling layer is tracked by a flattening procedure, which opens the values into a one-dimensional format to make the input for next layers.

2) Bi-LSTM block: The next segment of the method contains the BiLSTM module, which encompasses the BiLSTM network, a dropout layer, and a dense layer. To know the model and the performance of the BiLSTM network initially describes the one-directional LSTM network. The LSTM network is one of its kind of RNN, At this juncture they utilizes cyclic links in its inner layers which give short-term memory for the method and the capability to process thesequential data. Therefore, classical RNN have a problem known as the long-term dependencies issue, the poor memory of previous data as the neurons number increases. The LSTM network resolves this issue by storing relevant data by entire LSTM units in a type of conveyor belt called memory cell. Every LSTM unit includes a memory cell and 3 gates that regulate data flow by determining which data is forgotten and which one remains in the method. This is primary for learning the long-term dependencies through LSTM and resolving issues. In LSTM, three gates are there such as forget, input, and output gates. The parameters of the network used are Adam optimizer, softmax activation function, 50 epochs, and 64 batch sizes.

In forget gate, the sigmoid function is employed for data included in the present input X_t and past hidden state $(h-1)$. This process is represented as f_t returned the value among 0 and 1 which represents the percentage of data. The input gate considers data from present data and past hidden states and is passed by the second sigmoid function, transforming the data to values between 0 and 1. The similar data passed by the tanh function that supports to regulation of the network returns the value among -1 and 1. Next, the sigmoid result i_t is multiplied through $tanh$ result for determining which data is significant to keep. Now, there is enough data for executing the cell state. Initially, past cell state is multiplied through forget result. Next, result of input gate is included, updates cell state with new values considered relevance through network. The outcome of these two processes is provided as new cell state. At last, the output gate is provided, which determines value of following hidden state. Initially, data from present input and from past hidden state passed by third sigmoid function. Next, new cell state passes by tanh function. The mathematical formula for LSTM cell is given from Eq. (14) to Eq. (19),

$$f_t = \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big) \tag{14}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{15}$$

$$\hat{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{16}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \hat{C}_t \tag{17}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{18}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{19}$$

In Eq. (14) to Eq. (19), the σ describes the sigmoid function, the tanh describes the hyperbolic tangent function, x_t is input data, the h_t is a hidden state in time, the W_x and b_x are the weight matrix and bias vector. The Bi-LSTM network includes forward and backward LSTM networks. The forward LSTM utilizes the input sequence of values ranging from $t-k$ to t

when backward LSTM utilizes input sequence range from t to $t-k$. The result of the BiLSTM layer is attained by using Eq. (20),

$$Y_t = \sigma(\vec{h}_t, \overleftarrow{h}_t) \quad (20)$$

Here, the sigmoid function unified the results of unidirectional LSTM networks. The BiLSTM utilizes the dropout method for regulating the over-fitting.

IV. Experimental Results

The performance of developed technique is simulated by using python environment and used system configurations are i5 processor, 8 GB RAM and windows 10 (64 bit) The evaluation metrics utilized to assess performance are accuracy, recall, f1-score, precision and AUC/ROC. The True Negative (TN) and True Positive (TP) values represents the ability of classifier method for predicting the presence or absence of sepsis in patient. The False Negative (FN) and False Positive (FP) values represents incorrect predictions identified by methods. The accuracy represents ratio of actual positive observations to total number of positive instances. The recall executes whole fraction of positive instances. The f1-score calculates mean of precision and recall. The mathematical formula for evaluation metrics is given from Eq. (21) to (24),

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (21)$$

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (22)$$

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (23)$$

$$F1-score = \frac{2 \times Precision \times Recall}{Precision+Recall} \times 100 \quad (24)$$

In Table I, the feature selection algorithm CL-LBO is evaluated with different metrics on the MIMIC III dataset. The Honey Badger Optimization (HBO), Ant Colony Optimization (ACO), and Squirrel Search Algorithm (SSA) are the other feature selection algorithms considered to evaluate performance of CL-LBO algorithm. The developed feature selection algorithm reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared with other algorithms.

TABLE I. Performance of Feature Selection Algorithm

Feature selection algorithms	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
HBO	94.75	94.60	94.50	94.55	94.85
ACO	95.65	95.50	95.40	95.45	95.70
DSSA	96.85	96.70	96.60	96.65	96.90
CL-LBO	**99.85**	**99.60**	**99.50**	**99.55**	**99.95**

In Table II, the performance of classifier is evaluated using whole feature set with different metrics on the MIMIC III dataset. The RNN, Gated Recurrent Unit (GRU), and LSTM are the other classifiers considered to evaluate the performance of the CNN-LSTM network. The classifier with whole feature set reached 96.45% accuracy, 96.58% precision, 96.98% recall, 96.51% f1-score, and 96.45% AUC.

TABLE II. Performance of Classifier Using Whole Feature Set

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
RNN	93.50	93.60	93.70	93.65	93.40
GRU	94.80	94.85	94.90	94.87	94.75
LSTM	95.60	95.65	95.75	95.70	95.50
Proposed CNN-LSTM	**96.45**	**96.58**	**96.98**	**96.51**	**96.45**

In Table III, the performance of the classifier is evaluated using selected relevant features with different metrics on the MIMIC III dataset. The RNN, GRU, and LSTM are the other classifiers considered to evaluate the performance of the CNN-LSTM network. The classifier with selected relevant features reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared to other classifiers.

TABLE III. Performance of Classifier Using Selected Relevant Features

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
RNN	95.45	95.20	95.10	95.15	96.10
GRU	96.85	96.50	96.40	96.45	97.20
LSTM	97.75	97.60	97.50	97.55	98.10
Proposed CNN-LSTM	**99.85**	**99.60**	**99.50**	**99.55**	**99.95**

In Table IV, the performance of the activation function is evaluated with various metrics on the MIMIC III dataset. The Tanh, Rectified Linear Unit (ReLU) and sigmoid are other activation functions considered to evaluate the performance of the softmax function. The classifier with the softmax function reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared to other activation functions.

TABLE IV. Performance of Activation Function

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
Tanh	94.58	94.40	94.30	94.35	94.68
ReLU	95.75	95.60	95.50	95.55	95.89
Sigmoid	96.85	96.70	96.60	96.65	96.99
Softmax	**99.85**	**99.60**	**99.50**	**99.55**	**99.95**

In Table V, the performance of the optimizer is evaluated with various metrics on the MIMIC III dataset. The AdaGrad, Adamax, and Stochastic Gradient Descent (SGD) are other optimizers considered to evaluate the performance of the Adam optimizer. The classifier with Adam optimizer reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared to other optimizers.

TABLE V. Performance of Optimizer

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
AdaGrad	94.53	94.47	94.32	94.39	94.61
Adamax	95.76	95.63	95.54	95.58	95.82
SGD	96.87	96.73	96.65	96.69	96.91
Adam	**99.85**	**99.60**	**99.50**	**99.55**	**99.95**

In Fig. 3, the accuracy vs. epoch graph is represented for the developed classifier. The training accuracy, after some epochs, is close to 100% represents that the classifier has learned to classify the data accurately. The validation accuracy enhances the training process, represents that it stabilizes the model during training. There is only a small gap between training and validation accuracy representing that the model has better generalization capability and there is no overfitting. The validation accuracy stabilizes the model without any drop. In Fig. 4, the loss vs. epochs is represented for the developed classifier. The training loss continues to decline plateaus at less value, representing that the model has minimized errors in training data. The validation loss minimized in the initial phase represents that the method maximizes its performance on unseen data. After certain epochs, validation loss stabilizes the model. There is less gap between training and validation loss, which represents that the model is not overfitting.

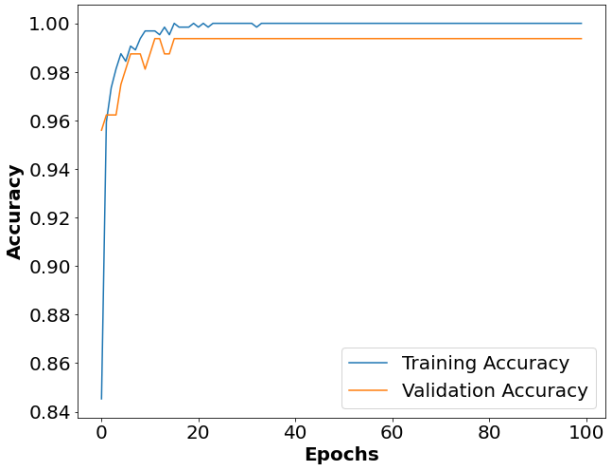

Fig. 3. Accuracy vs. Epochs.

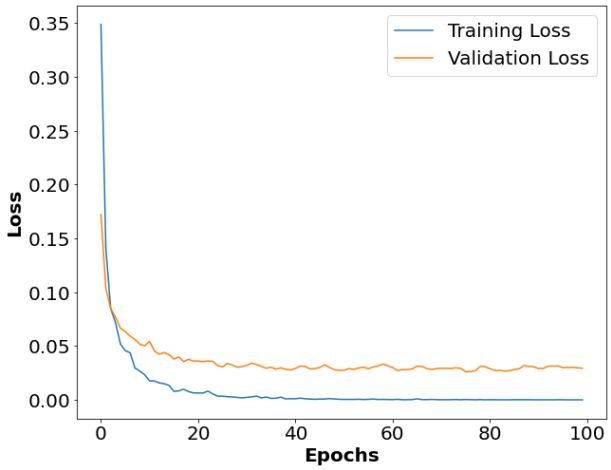

Fig. 4. Loss vs. Epochs.

A. AUC/ROC Curve

By using the AUC/ROC curve, Fig. 5 represents the connection between True Positive Rate (TPR) and False Positive Rate (FPR) by utilizing the AUC/ROC function. The ROC determines the ability of classifier for differentiating their classes. While AUC is substantial, the prediction of the method is correct.

B. Comparative Analysis

In Table VI, performance of the implemented technique is compared to existing techniques like LSTM [18], Ensemble ML [19], and CNN-Bi-LSTM [23] with different metrics on MIMIC III dataset. The CNN extracts the significant features from relevant features, focused on spatial-based relationships. Then, the Bi-LSTM layer captured the sequential dependencies and temporal relationships in patient histories that are essential to understand the treatment results. CL-LBO integrates the circle chaotic map and Levy flight process in traditional LBO to select relevant features for classification. The proposed technique reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared to existing techniques.

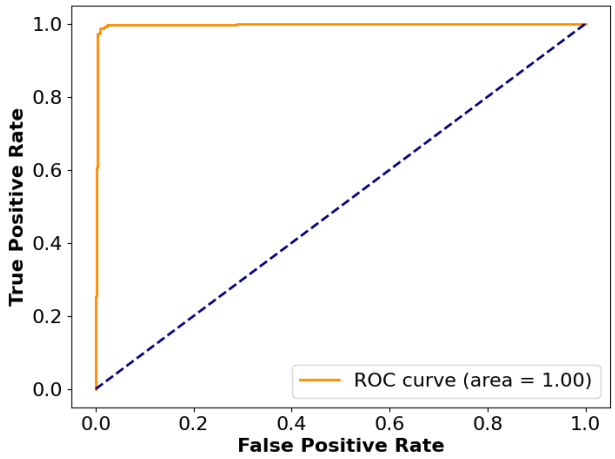

Fig. 5. ROC curve.

TABLE VI. COMPARATIVE ANALYSIS

Methods	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
LSTM [18]	MIMIC III	-	77.00	79.87	78.24	85.01
Ensemble ML [19]		92.80	-	-	-	83.00
CNN-Bi-LSTM [23]		99.15	99.16	99.15	98.85	-
Proposed technique		**99.85**	**99.60**	**99.50**	**99.55**	**99.95**

C. Discussion

The outcomes of the proposed technique are evaluated with the MIMIC III dataset using various evaluation metrics. The developed CL-LBO feature selection algorithm is evaluated with different optimization algorithms of HBO, ACO, and DSSA. The developed classifier is evaluated with default features and with selected relevant features. Additionally, the performance of the classifier is evaluated with an activation function and different optimizers. Moreover, the performance of the developed technique is compared with LSTM [18],

Ensemble ML [19], and CNN-Bi-LSTM [23] with different metrics on the MIMIC III dataset. These existing algorithms have drawbacks such as overfitting issues occurring during the process because of the irrelevant features present in the data fit, dimensionality issues, struggled to interpret the pattern of features and the missing values presented in the data caused an overfitting problem. To mitigate these drawbacks, the CNN-Bi-LSTM method for personalized treatment analysis uses EHR data. The CNN extracts the significant features from relevant features, focused on spatial-based relationships. Then, the Bi-LSTM layer captured the sequential dependencies and temporal relationships in patient histories that are essential to understand the treatment results. The SMOTE technique is employed in the pre-processing phase to balance the classes in data. The CL-LBO integrates the circle chaotic map and Levy flight process in traditional LBO to select relevant features for classification. This process improves the process of personalized treatment with high classification accuracy. By using this technique, in this article, the developed technique reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared to existing techniques.

V. CONCLUSION

The traditional mortality risk prediction techniques effectively extract the data in longitudinal EHRs that ignore the difficult relationship and interactions among variables and time dependency in longitudinal records. In the proposed work, the CNN-Bi-LSTM method is developed for personalized treatment analysis using EHR data. The MIMIC III dataset is used in this article and the data is balanced by using SMOTE and the balanced data is encoded by using the one-hot encoding technique. Then, the relevant features from pre-processed data are selected by using the developed CL-LBO algorithm. Here, the circle chaotic map and levy flight process are integrated with the traditional LBO algorithm to enhance the search capability and convergence rate of the algorithm. Then, the CNN extracts significant features from relevant features, focused on spatial-based relationships. Then, the Bi-LSTM layer captured the sequential dependencies and temporal relationships in patient histories that are essential to understand the treatment results. The proposed method reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared to LSTM. In the future, different DL-based will be used to further enhance the process of personalized treatment analysis.

REFERENCES

[1] W. Wang, P. Mohseni, K. L. Kilgore, and L. Najafizadeh, "PulseDB: A large, cleaned dataset based on MIMIC-III and VitalDB for benchmarking cuff-less blood pressure estimation methods," Front. Digit. Heal., vol. 4, p.1090854, February 2023, doi: 10.3389/fdgth.2022.1090854.

[2] R. Zhang et al., "Independent effects of the triglyceride-glucose index on all-cause mortality in critically ill patients with coronary heart disease: analysis of the MIMIC-III database," Cardiovasc. Diabetol., vol. 22, no. 1, p.10, January 2023, doi: 10.1186/s12933-023-01737-3.

[3] J. Xu, H. Cai, and X. Zheng, "Timing of vasopressin initiation and mortality in patients with septic shock: analysis of the MIMIC-III and MIMIC-IV databases," BMC Infect. Dis., vol. 23, p.199, April 2023, doi: 10.1186/s12879-023-08147-6.

[4] S. Peng, J. Peng, L. Yang, and W. Ke, "Relationship between serum sodium levels and all-cause mortality in congestive heart failure patients: A retrospective cohort study based on the Mimic-III database," Front. Cardiovasc. Med., vol. 9, p.1082845, January 2023, doi: 10.3389/fcvm.2022.1082845.

[5] W. Liao and J. Voldman, "A Multidatabase ExTRaction PipEline (METRE) for facile cross validation in critical care research," J. Biomed. Inform., vol. 141, p. 104356, May 2023, doi: 10.1016/j.jbi.2023.104356.

[6] Y. Hakverdi et al., "Enhancing ICU Management and Addressing Challenges in Türkiye Through AI-Powered Patient Classification and Increased Usability With ICU Placement Software," IEEE Access, vol. 12, pp. 146121–146136, 2024, doi: 10.1109/access.2024.3426919.

[7] J. Chen, T. Di Qi, J. Vu, and Y. Wen, "A deep learning approach for inpatient length of stay and mortality prediction," J. Biomed. Inform., vol. 147, p. 104526, November 2023, doi: 10.1016/j.jbi.2023.104526.

[8] S. Wei et al., "Machine learning-based prediction model of acute kidney injury in patients with acute respiratory distress syndrome," BMC Pulm. Med., vol. 23, no. 1, p.370, October 2023, doi: 10.1186/s12890-023-02663-6.

[9] N. Ashrafi, Y. Liu, X. Xu, Y. Wang, Z. Zhao, and M. Pishgar, "Deep learning model utilization for mortality prediction in mechanically ventilated ICU patients," Informatics Med. Unlocked, vol. 49, p. 101562, 2024, doi: 10.1016/j.imu.2024.101562.

[10] L. Liu, O. Perez-Concha, A. Nguyen, V. Bennett, and L. Jorm, "Automated ICD coding using extreme multi-label long text transformer-based models," Artif. Intell. Med., vol. 144, p. 102662, October 2023, doi: 10.1016/j.artmed.2023.102662.

[11] M. Bernardini, A. Doinychko, L. Romeo, E. Frontoni, and M.-R. Amini, "A novel missing data imputation approach based on clinical conditional Generative Adversarial Networks applied to EHR datasets," Comput. Biol. Med., vol. 163, p. 107188, September 2023, doi: 10.1016/j.compbiomed.2023.107188.

[12] S. S. Samy, S. Karthick, M. Ghosal, S. Singh, J. S. Sudarsan, and S. Nithiyanantham, "Adoption of machine learning algorithm for predicting the length of stay of patients (construction workers) during COVID pandemic," Int. J. Inf. Technol., vol. 15, no. 5, pp. 2613–2621, June 2023, doi: 10.1007/s41870-023-01296-6.

[13] M. Fathima Begum and S. Narayan, "A pattern mixture model with long short-term memory network for acute kidney injury prediction,"J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 4, pp. 172–182, April 2023, doi: 10.1016/j.jksuci.2023.03.007.

[14] C. Yu and Q. Huang, "Towards more efficient and robust evaluation of sepsis treatment with deep reinforcement learning," BMC Med. Inform. Decis. Mak., vol. 23, no. 1, p.43, March 2023, doi: 10.1186/s12911-023-02126-2.

[15] A. Ahmed, X. Zeng, R. Xi, M. Hou, and S. A. Shah, "MED-Prompt: A novel prompt engineering framework for medicine prediction on free-text clinical notes," J. King Saud Univ. - Comput. Inf. Sci., vol. 36, no. 2, p. 101933, February 2024, doi: 10.1016/j.jksuci.2024.101933.

[16] X. Li, Y. Zhang, X. Li, H. Wei, and M. Lu, "DGCL: Distance-wise and Graph Contrastive Learning for medication recommendation," J. Biomed. Inform., vol. 139, p. 104301, March 2023, doi: 10.1016/j.jbi.2023.104301.

[17] H. Dong et al., "Ontology-driven and weakly supervised rare disease identification from clinical notes," BMC Med. Inform. Decis. Mak., vol. 23, no. 1, p.86, May 2023, doi: 10.1186/s12911-023-02181-9.

[18] F. Yang, J. Zhang, W. Chen, Y. Lai, Y. Wang, and Q. Zou, "DeepMPM: a mortality risk prediction model using longitudinal EHR data," BMC Bioinformatics, vol. 23, no. 1, p.423, October 2022, doi: 10.1186/s12859-022-04975-6.

[19] N. Tasnim, S. Al Al Mamun, M. S. Shahidul Islam, M. S. Kaiser, and M. Mahmud, "Explainable Mortality Prediction Model for Congestive Heart Failure with Nature-Based Feature Selection Method," Appl. Sci., vol. 13, no. 10, p. 6138, May 2023, doi: 10.3390/app13106138.

[20] M. Bampa, I. Miliou, B. Jovanovic, and P. Papapetrou, "M-ClustEHR: A multimodal clustering approach for electronic health records," Artif. Intell. Med., vol. 154, p. 102905, August 2024, doi: 10.1016/j.artmed.2024.102905.

[21] B. Alsinglawi et al., "An explainable machine learning framework for lung cancer hospital length of stay prediction," Sci. Rep., vol. 12, no. 1, p.607, January 2022, doi: 10.1038/s41598-021-04608-7.

[22] S. Niu, J. Ma, L. Bai, Z. Wang, L. Guo, and X. Yang, "EHR-KnowGen: Knowledge-enhanced multimodal learning for disease diagnosis generation," Inf. Fusion, vol. 102, p. 102069, February 2024, doi: 10.1016/j.inffus.2023.102069.

[23] S. Sakri et al., "Sepsis Prediction Using CNNBDLSTM and Temporal Derivatives Feature Extraction in the IoT Medical Environment," Comput. Mater. & Contin., vol. 79, no. 1, pp. 1157–1185, 2024, doi: 10.32604/cmc.2024.048051.

[24] S. R. Khope and S. Elias, "Simplified & Novel Predictive Model using Feature Engineering over MIMIC-III Dataset," Procedia Comput. Sci., vol. 218, pp. 1968–1976, 2023, doi: 10.1016/j.procs.2023.01.173.

[25] C. Liu et al., "Early prediction of MODS interventions in the intensive care unit using machine learning," J. Big Data, vol. 10, no. 1, p.55, May 2023, doi: 10.1186/s40537-023-00719-2.

[26] V. K. Chauhan, A. Thakur, O. O'Donoghue, O. Rohanian, S. Molaei, and D. A. Clifton, "Continuous patient state attention model for addressing irregularity in electronic health records," BMC Med. Inform. Decis. Mak., vol. 24, no. 1, p.117, May 2024, doi: 10.1186/s12911-024-02514-2.

[27] H.-J. Lee et al., "StrokeClassifier: ischemic stroke etiology classification by ensemble consensus modeling using electronic health records," npj Digit. Med., vol. 7, no. 1, p.130, May 2024, doi: 10.1038/s41746-024-01120-w.

[28] A. Johnson, T. Pollard, and R. Mark, "MIMIC-III Clinical Database." PhysioNet, 2023, doi: 10.13026/C2XW26.

[29] J. He, Y. Hao, and X. Wang, "An Interpretable Aid Decision-Making Model for Flag State Control Ship Detention Based on SMOTE and XGBoost,"J. Mar. Sci. Eng., vol. 9, no. 2, p. 156, February 2021, doi: 10.3390/jmse9020156.