The Bioscan 19(2): S.I (1), 288-293, 2024 www.thebioscan.com
AN INTERNATIONAL QUARTERLY JOURNAL OF LIFE SCIENCES

Integrating Genetic Algorithm and Leverage the Slap Swarm Algorithm to Optimize the LSTM Model to Predict Cardiovascular Disease

Mr. A. Rameshbabu [1], Dr G.N.R. Prasad [2], Dr Ravi Kumar M [3], V. Usha [4],

Dr. A. Azhagu Jaisudhan Pazhani [5], Dr. G. Charles Babu [6]

1 Department of Electronics & Communication Engineering, Ramco Institute of Technology, Rajapalayam, India. eceramesh babu@gmail.com

2 Department of MCA, Chaitanya Bharathi Institute of Technology, Hyderabad, India. gnrp@cbit.ac.in

3 Department of Electronics & Communication Engineering, Cambridge Institute of Technology, Bengaluru, India. ravikumar.ece@cambridge.edu.in

4 Department of Computer science and engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, v.ushavelusamy@gmail.com

5 Department of Electronics & Communication Engineering, Ramco Institute of Technology, Rajapalayam, Tamilnadu, Mail: alagujaisudhan@gmail.com

6 Department of Computer science and engineering, Gokaraju Rangaraju Institute of Engineering and Technology, Bachupally, Hyderabad, India. charles1624@grietcollege.com

DOI: https://doi.org/10.63001/tbs.2024.v19.i02.S.I(1).pp288-293

KEYWORDS

Genetic Algorithm, Salp Swarm Algorithm, Cardiovascular Illnesses, Clinical Practice, Diagnostic Models

Received on:

08-08-2024

Accepted on:

25-11-2024

ABSTRACT

To efficiently avoid cardiovascular ailments, that are ordinary and a major community health question, early discovery is essential. Problems accompanying network arrangement and acting decay stretch to affect the veracity of demonstrative models, even though skilled are many of the ruling class. This research suggests the OCI-LSTM model, which signifies well-organised and enhanced long-term working memory, as a forceful answer. The LSTM's network design is optimised using the Genetic Algorithm, and superfluous traits are accurately erased utilizing the Salp Swarm Algorithm. The model's efficiency is habitual by confirmation measures, which involve the F1 score, particularity, sense, and veracity. The OCI-LSTM outperforms both the in research comparing Deep Neural System and the Deep Belief System, the outcomes showed that the former achieved 98.12% better accuracy. These advancements have transformed the OCI-LSTM into a paradigm for the efficient, accurate, and early detection of cardiovascular diseases. To ensure a smooth transition into clinical practice, future studies may investigate real-world applications and methods for further improvement.

INTRODUCTION

The main cause of the recent increase in deaths from cardiovascular disease is problems with how efficiently the heart can distribute blood throughout the human body, which leads to disturbances in blood circulation. Heart disease (CVD) is the most harmful condition to human health within the range of heart-related conditions. CVD is now a major cause of elevated death rates due to its rising incidence, which poses significant difficulties to the worldwide healthcare sector. Surveys show that CVD has killed as many as four out of ten people, impacting around a million people.

Many studies have utilised various characteristics, such as sex, age, fasting blood sugar, discomfort in the chest (including its type and location), blood sugar level, smoking status, depression, electrocardiogram, results of relaxing electrocardiogram, gradient, old spike, heart status, inadequate diet, cholesterol levels, weight gain, family history, intake of alcohol, hypertension, and physical inactivity. Invasive procedures based on a patient's health record & an analytical statement by a health expert have typically been used for CVD prediction. Not to

mention how difficult and expensive it is. To tackle these difficulties, non-invasive technologies like clinical decision-making assistance models that use Deep Learning (DL) and Machine Learning (ML) methodologies are crucial.

Convolutional neural networks represented as CNN-LSTM, was used to identify corona virus 19. Three different X-ray picture types are used in this combination to predict illness, and LSTM is used as the classifier to differentiate between different COVID-19 instances. It successfully addresses the problem of overfitting and guarantees improved outcomes for picture datasets of different sizes and resolutions.

Several fields have benefited greatly from Deep Learning's applications, such as medical imaging, illness monitoring, drug development, protein structure analysis, and evaluating the severity and infectiousness of the COVID-19 virus. Machine learning, deep knowledge, and mathematical wisdom are just any examples of up-to-date electronics that are helping in the battle against hazardous ailments.

Resolving questions that have damaged prognosis models' accomplishment is the main aim of the submitted OCI-LSTM model. Overfitting and underfitting are generated by dossier preparation, that is the primary question. Optimising the arrangement of the network model is another question. Even minor features of the preparation dossier power cause the model to overfit, that produces weak effects if used to test facts. Underfitting happens when the model learns poorly, bearing inferior consequences from both preparation and testing dossier. The main causes of these questions are the network model's vulgar arrangement and design, in addition to the addition of unnecessary features. Thus, the computational cost & the forecast opportunity for CVD are raised by these questions. In order to answer that, the SSA is executed to kill repetitious or boisterous features, that aids ineffectively recognizing best choice traits. Additionally, a better LSTM for categorization is submitted, accompanying the network design optimised utilizing the Genetic Algorithm (GA). By selecting the ideal period window capacity, providing an ideal resolution, and reconstructing model depiction, the GA regulates the model.

At last, we set the model through allure paces by attending trials and attracting four key acting measures into the report. Training and experiments are administered utilizing the Indian dataset, which is frequently used in studies on coronary thrombosis and maybe about the public warehouse.

2. The suggested OCI-LSTM method:

Researchers determined a novel approach created to overcome network generalisation issues, including overfitting and underfitting, and increase the accuracy of CVD prediction. It also addresses optimisation and configuration-related problems, such as figuring out the best network setup. The min-max scaling technique is used for preprocessing at the start of the procedure. The SSA is then used to choose the top qualities. These improved characteristics are then put into effect with OCI-LSTM, which successfully fixes the aforementioned problems and raises the degree of prediction accuracy.

2.1. Normalising Features via Min-Max Scaling:

As part of the preprocessing step, the minimum - maximum scaling method is used to normalise missing values and unnecessary data. Factors included in the dataset include patient age, sex, kind of chest pain, blood pressure at rest and while fasting, maximal heart rate, exercise-induced chest discomfort, number of main arteries, and elevation of the ST segment. Thorough data gathering for the repository is facilitated by routine and methodical monitoring. Problems with missing data values, patients interfering, and technological issues influencing illness analysis emerge during data gathering. In the end, only relevant data remains after this method has removed any irrelevant or noisy data.

It was found that seven of the dataset's 340 occurrences have missing values after some analysis. The normalisation procedure is used to guarantee standardised outcomes for simple interpretation during model training, and Equation (1) is used to get the standardised output.

$$\text{Standardised Output} = \frac{Y_{Value} - mean}{SD} \times 100 \qquad (1)$$

where Y_{value}Is the highest value of the cardiac data, and SD is the standard deviation.

Equation (2) is used to find the dataset mean.

$$\text{Mean} = \frac{\sum_{i=1}^{n} y}{n} \qquad (2)$$

Equation (3) is used to figure out SD.

$$\text{SD} = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(y_i - Mean)^2} \qquad (3)$$

Where the standard deviation is computed from a total of N samples.

The scaling procedure, however, is not well-suited to the categorical characteristics. Consequently, the values are adjusted into an interval of 0–1 using the min-max technique. This modification facilitates the model's data interpretation during training.

The following is how the data normalisation is carried out:

$$N' = \frac{Y_{value} - Y_{\min\_value}}{Y_{\max\_value} - Y_{\min\_value}} \qquad (4)$$

Normalised data is indicated by N′, and a specific data point of each instance is indicated by Y_{value}, The dataset's lowest value is represented by $Y_{\min\_value}$, The dataset's highest value is represented by $Y_{\max\_value}$.

To reduce the risk factors linked to CVD in disease prediction, this research expands its analysis by evaluating several metrics, including variance, maximum, minimum, correlation, and energy. At the outset, we replace features that don't seem to be adding any value with brand-new ones. There is a risk of overfitting since the dataset frequently includes a large amount of patient data, and while certain characteristics may be useful for illness prediction, others may not be. This is why the study uses optimisation and feature set reduction to improve illness detection. To identify the most relevant optimised properties from the original dataset, the Salp Swarm approach is utilised, as explained in the following section.

2.2. Finding the Best Subset Feature with Salp Swarm:

To improve the model's learning process by getting rid of useless attributes, the SSA is used for attribute selection. As a means of population selection, the SSA makes use of the swarming process seen in salps, a creature native to the ocean. An oceanic phenomenon known as a salp swarm is formed when a single salp, the head salp, is followed closely by several other salps. An n-dimensional search point denotes a salp position, where 'n' is the overall quantity of identifiers in a particular issue. All in all, there are three stages of feature optimisation: 1. populating the system with starting values, 2. modifying the position of the leader, and 3. modifying the position of the follower. The salp swarm's clustering process is mirrored by these processes. A thorough explanation of the SSA's operational premise is provided in the parts that follow.

2.2.1. Population Initialisation:

The population initialisation is done by using the S × D Geometric workplace, where S indicates swarming size and D indicates space dimension. Designate the diet present in the location as F_D. Nd is allocated as $F_D = \left(F_{D_1}, F_{D_2}, \ldots\ldots, F_{D_d}\right)^T$. $P_n$represents the location of each salp where $P_n = (P_{n1}, P_{n2}, \ldots\ldots, P_{nd})^T$. U_b and $L_b$denotes the upper and lower bound respectively, where U_b takes the values as $U_b = \left(U_{b_1}, U_{b2}, \ldots\ldots, U_{bd}\right)^T$, and L_b as $L_b = \left(L_{b_1}, L_{b2}, \ldots\ldots, L_{bd}\right)^T$.

Equation (5) is used to calculate the population's random initialisation.

$$Y_{S\times D} = rand(S \times D) \times (U_b - L_b) + L_b \times ones(S \times D) \qquad (5)$$

Both the leader and the following $Y_{1,d}$ and $Y_{K,d}$ represents the population's condition in the dth dimension, while K takes the values from 2 to N.

2.2.2. Revising Leading Part:

Finding food within the area is the leader's responsibility in a salp swarm. Additionally, it has to direct the whole group while they look for food. The leader's position must be updated, and Equation (6) is used to do this.

$$Y_{1,d} = fd_d + r_1((U_{bd} - L_{bd})r_2 + L_{bd}) \quad (6)$$

Global exploration is carried out by the algorithm if r_1 is bigger than one. It concentrates on resident investigation to identify a correct estimate rate if r_1 is smaller than one. For the algorithm's first iteration to do a global search & later increase the precision in future repetitions, the initial value of r_1 should be between 2 and 0. Equation (7) is used to compute the convergence factor.

$$r_1 = 2e^{-\left[\frac{4i}{i_{max}}\right]^2} \quad (7)$$

The present iteration is represented as i and the highest number of iterations is represented as i_{max}.

2.2.3. Revising the Position of the Follower:

Throughout the SSA, the adherents engage in a sequence of coordinated motions instead of arbitrary actions. To ascertain the movement of followers, it is essential to evaluate many key factors, including their beginning location, velocity, and acceleration.

Equation (8) is used to compute the motion distance, which is determined by using Newton's equation of motion.

$$\text{Motion Distance} = \frac{1}{2}ai^2 + S_f i \quad (8)$$

Where a is the acceleration and S_f Is the followers' speed.

Equation (9) is used to determine the acceleration of the followers.

$$a = \left[\frac{S_{final} - S_f}{t}\right] \quad (9)$$

Equation (10) may be used to find the salp's movement speed.

$$S_{final} = \frac{\left[Y^i_{K-1,d} - Y^i_{K,d}\right]}{t} \quad (10)$$

Finally, the Motion distance is represented as, $\text{Motion Distance} = \frac{1}{2}\left[Y^i_{K-1,d} - Y^i_{K,d}\right]$

Based on this the followers' position is revised in equation (11).

$$Y^i_{K,d} = Y^i_{K,d} + \text{Motion Distance} = \frac{1}{2}\left[Y^i_{K-1,d} - Y^i_{K,d}\right] \quad (11)$$

2.3. Optimisation using Genetic Algorithm:

The fittest members of each generation are chosen to determine the GA's optimal solutions. Several operators choose eligible individuals from the current generation as part of the GA's core procedure.

The involvement of people in fitness-based generational selection is made possible by the selection operator. By assessing the suitability of the present generation, it determines the next population level. Two popular GA selection operators are the Stochastic Universal Sample (SUS) & the Roulettette Wheel (RW). The RW determines the odds of selection for each individual independently. To get an individual's proportional fitness selection, use equation (12):

$$P_i = \frac{F_i}{\sum_{K=1}^{N} F_K} \quad (12)$$

Here F_i represent individual fitness, P_i denotes the likelihood of an individual & N indicates overall individuals in the population.

Replacement Operator: This function is essential to the propagation process and makes it easier for members to be passed down from one generation to the next:

$$F_i = W_1 \times age + W_2 \times restecg + W_3 \times maxhr \ldots. \quad (13)$$

Where W_1, W_2, $W_3 Are$ values given to various health metrics.

Recombination operator: This operator is used to swap substrings of two distinct members of the same generation by the recombination operator, which makes use of the intersection idea. There are three common methods for the recombination operator: uniform, two-point, and single-point crossings.

Genetic Drift Operator: This operator is charge of altering current generation members' genes to produce the next generation. There are many ways to carry out mutations, such as Gaussian, border, non-uniform, and uniform mutations. Among these techniques, the Gaussian operators is often used, which adds random values from a typical distribution to the chosen gene. Consider if $Y \in (w, z)$, a selected gene is utilized for doing the mutation progression. Thereby Y'Is computed in equation (14).

$$Y' = min[max(M_o(y, r), w), z] \quad (14)$$

Here $M_o(y, r)$indicates mutation operation, and r rate of mutation depending on time duration.

2.4. OCI-LSTM Model:

To choose the best time frame for the LSTM units, the researcher provides the OCI-LSTM model, which combines it with GA. An important benefit of the LSTM is that it improves the model's performance by learning the optimal time frame from historical data. Choosing the right window size is critical; if it's too big, the model will get overfitted with training data, and if it's too little, the network will miss important information. Figure 1 illustrates an outline of the overall framework.

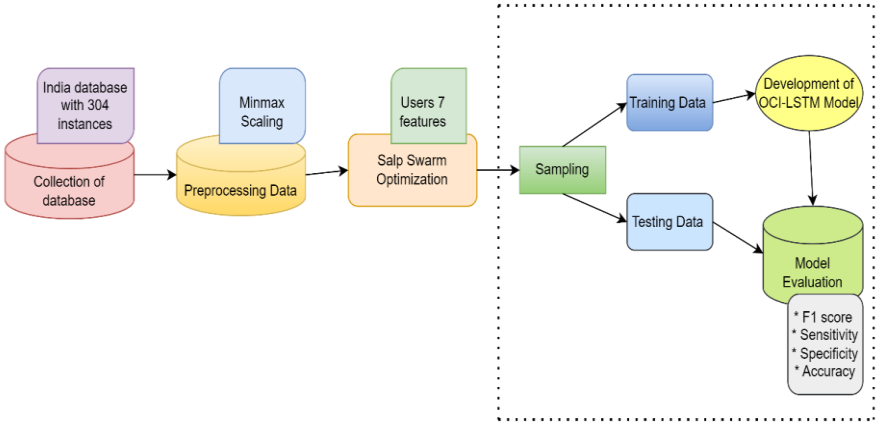


Figure 1, depicts the structure of the OCI-LSTM model.

There are two stages in the OCI-LSTM model. Parameters are appropriately established in the initial step. There are two hidden layers and an input layer in the network. The GA makes sure there is an ideal number of hidden neurones in the hidden layers. The OCI-LSTM network uses two activation functions. The linear function is used for the output nodes by considering the origin of the CVD prediction issue. The sigmoid function is used at the hidden and input nodes to alter input values from -1 to 1. The Adam optimiser, which is renowned for its computing efficiency, is used to modify the network weights after they are first assigned random values. The GA, an evolutionary-based search technique, is used to investigate the OCI-LSTM network's architectural aspects and identify the ideal window size.

Phase two of the network design involves assessing the GA's fitness. For this study, we use a variety of hidden layer LSTM units and apply different window widths to the OCI-LSTM. Before utilising the operators to explore the two-dimensional space, the populations, which were originally given random values, go through an initialisation phase.

In this study, the chromosomes are stored as binary bits that indicate the LSTM cell numbers and the size of the temporal frame. After that, genetic operators look for the optimal answer and assess each one using the fitness approach. In this study, the fitness function is MSE. The ideal solution is the one with the lowest MSE value that the architectural elements produce. The calculated optimum solution is used for the OCI-LSTM model in-

case the termination criteria are met. If not, the whole genetic process is again carried out till the need is satisfied. During testing, parameters like the population size, crossover, and mutation are changed to increase the model's fitness and provide better outcomes. To achieve the termination criterion in this research, the OCI-LSTM runs for ten generations with 0.8 as overlap variable value & 0.16 as mutation value.

Researchers make use of the India dataset, which is accessible to the public via online repositories. There is a grand total of 77 characteristics in the dataset. However, most researchers only consider 14 of the 77 traits when they look for signs of heart disease. There are 304 cases in the dataset, and 7 of them have missing values. Researchers start by eliminating the outliers in the dataset in the preprocessing step of the suggested OCI-LSTM procedure. Consequently, 7 cases are eliminated, leaving 298 occurrences for further analysis. Table 1 depicts the 12 features from the India dataset.

To improve the model's performance, it is essential to use the SSA to choose the most essential characteristics. The six attributes like A_1, A_8, A_{10}, A_{11} *and* A_{12}Were chosen as the best ones according to the SSA. Considering the Indian Heart Disease dataset, each of these characteristics is highly predictive of CVD.

Table 1, illustrates the features of the India dataset.

S.no	Indicate	Category	Data [Binary(B)/Nominal(No)/Numeric(Nu)]
1	A_1	Age	Nu
2	A_2	Gender	B
3	A_3	Hp_type	No
4	A_4	Rest_bp	Nu
5	A_5	Ser_Choles	Nu
6	A_6	Fasting_bp	B
7	A_7	Relax_ecg	No
8	A_8	Maxhr Rate	Nu
9	A_9	Exia	B
10	A_{10}	Slopst	No
11	A_{11}	Nvesl	No
12	A_{12}	Thalas	No

Together, these carefully chosen characteristics provide vital details on the patient's age, metabolic profile, resting and exercising heart performance, and symptom presence—all of which are critical components in estimating the likelihood of CVD. By streamlining the attribute selection process, the SSA reduces overfitting and improves overall performance while guaranteeing that the selected qualities make a substantial contribution to the model's predicted accuracy.

Information regarding each patient's multiple visits, the duration of the interval between each visit, whether or not the target variable is a cardiac patient, and the number of actions that the LSTM system should take into account when making predictions are all part of the temporal features of the data that are processed by the proposed network.

A multi-class variable with a possible value between 0 and 4 is the dataset's class label. Values ranging from 1 to 4 indicate different phases of CVD presence, with 0 indicating no CVD at all. This research uses the binary class label system, where labels 0 and 1 indicate the presence and absence of CVD, respectively, from a multi-class label.

After implementing this change, 165 out of 298 cases in the dataset were found to be healthy participants or sick (labelled 0). Equation (16) represents the price of a GA-SSA algorithm.

$$Price(Y) = W_1 \times P_{GA}(Y) + W_2 \times P_{SSA}(Y) \quad (15)$$

Where weights W_1 and W_2 regulate how much each algorithm contributes to the total cost; Y indicates the solution, which could comprise genetic parameters and salps positions; $P_{GA}(Y)$ is the price of GA component and $P_{SSA}(Y)$ is represent the price of Salp Swarm the Algorithm component.

3. Results and Discussion:

To improve CVD prediction, a new method is suggested, the OCI-LSTM, which successfully reduces the effects of overfitting and underfitting. After evaluating other optimisation algorithms, the SSA proved to be the most effective in solving generalisation problems, hence it was chosen to pick a relevant feature subset for the technique.

By recognising temporal designs, like the ideal time window size, and using GA to find the finite LSTM units, the OCI-LSTM model overcomes network setup difficulties. The model is improved iteratively by integrating a local examination by GA, which guarantees that the GA finds the best-concealed layers in the LSTM and produces the best OCI-LSTM design.

Thorough comparisons with traditional models, DNNs, and DBNs allow for an in-depth examination of the OCI-LSTM. The findings demonstrate the best convergence and maximum accuracy rate.

Using a tiny portion of Indian statistics on cardiovascular disease with a modest dataset of variables, the suggested OCI-LSTM remarkably produces excellent results. This demonstrates the model's efficiency in predicting CVD and underlines its ability to beneficial in actual situations.

Two-way partitioning of the dataset into testing & training sets was executed. The test set was utilised to assess the classifier's performance, while the training set made model training easier. Every partition functioned as a test fold to guarantee thorough training and testing over 11 folds. Important criteria such as preciseness, sensitiveness, accuracy, and F1 score were used in the performance assessment. Equations (16) through (19) were used to compute these measurements, accordingly. A thorough evaluation of the OCI-LSTM model's performance was made possible by the findings, which were aggregated over the 11 folds.

$$\text{Accuracy} = (C_N + C_P) \div (C_N + C_P + IC_N + IC_P) \quad (16)$$

$$\text{Sensitiveness} = (C_P) \div (IC_N + C_P) \quad (17)$$

$$\text{Preciseness} = (C_P) \div (IC_P + C_P) \quad (18)$$

$$\text{F1 score} = [2 * \text{Preciseness} * \text{Sensitiveness}] \div [\text{Preciseness} + \text{Sensitiveness}] \quad (19)$$

Here C_N indicates Correct Negatives, C_P indicates Correct Positives, IC_N represent Incorrect Negatives and IC_P represent Incorrect Positives.

Splitting the whole Indian dataset into two halves is the first step of the experiment. The researcher divided the 298 cases into two parts: the training set has 208 instances and the testing set contains 90 instances. In Figure 2 the breakdown of the occurrences in the training set and the testing set is provided.

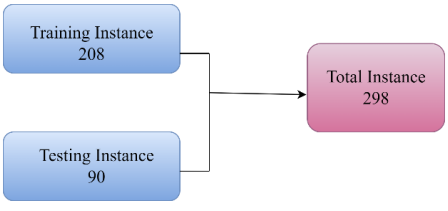
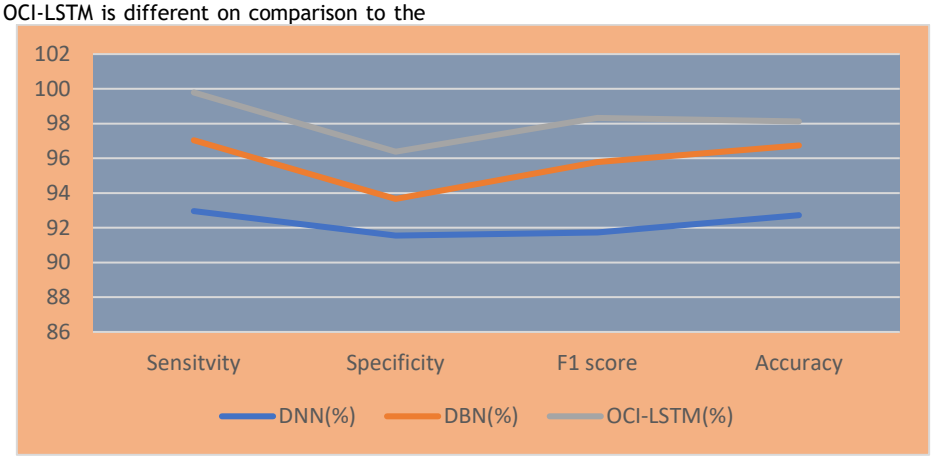


Figure 2, shows the breakdown of the occurrences.

In graph 1, the researcher can see how the three models fared in comparison to one another. Graph 1's comparative study shows that performance of OCI-LSTM is different on comparison to the other two networks' models. The OCI-LSTM yields better results for all four metrics on comparison of DNN and DBN.



Graph 1, illustrates the overall performance of all models.

Table 2 makes it clear that, in comparison to other well-known feature optimisation techniques, the categorical precision of the OCILSTM has increased when the SSA is used.

Table 2, shows the precision of the OCILSTM.

Algorithm	Attributes	Precision [DNN in percentage]	Precision [DBN in percentage]	Precision [OCI-LSTM in percentage]
GA	to7(A_2 *to* A_{812})	82.15	93.63	95.52
SSA	6(A_1, A_8 *to* A_{12})	92.73	96.74	98.12
CFS	to(A_6 *to* A_{12})	82.83	88.33	89.46
Chi	totoA_1 *to* A_5 *and* A_8*to* A_{12})	88.89	91.18	92.47
LASSO	to7(A_2, A_3, A_4 *and* A_7*to*A_{10})	85.52	92.47	95.8
RT	to6(A_3, A_4 *and* A_9 *to*A_{12})	91.58	95.62	97.83

A feature optimisation phase employing the SSA is the starting point for the OCI-LSTM. To improve the model's performance, the strategic feature selection procedure seeks to discover the most important qualities. To provide vital insights for the prediction of cardiovascular illness, seven optimised attributes (A_3, A_7, A_8, A_9, A_{10}, A_{11} *and* A_{12}) were chosen from a pool of twelve features based on their significant correlations with the objective variable. Feature optimisation makes use of the SSA, which is built to avoid overfitting and other network generalisation concerns. When a model becomes too used to the noise in its training data and starts to apply that knowledge to new data, it's overfitting. The OCI-LSTM minimises the risk of overfitting while ensuring excellent results for both testing and practice surfaces. This is achieved by selecting the optimal characteristics.

The usefulness of OCI-LSTM is tested in three thorough tests. The first trial involves using the OCI-LSTM on the best traits that were picked. Using the same best features, the second experiment compares the DNN and DBN models and writes down the results. Training OCI-LSTM, DBN, and DNN is done on the original, feature-rich set for the final test. The OCI-LSTM is tested further using a number of different feature selection methods. Furthermore, these tests show that the OCI-LSTM can help predict CVD and give a thorough analysis. The next step for OCI-LSTM was to finetune the LSTM units' time window sizes using a GA. Using this combination, the model adapts to the time-related aspects of the data so that it doesn't become too perfect while still gathering important data. Using a GA-based network setup process makes the model work better overall.

Overall, the OCI-LSTM outperforms the other classifiers and conventional approaches, particularly when combined with GA and Salp Swarm. After comparing the OCI-LSTM to the DNN and DBN, as well as models that include all of the features, it becomes clear that the former has superior predictive capabilities. The employment of algorithms based on evolution and the careful selection of optimal attributes contributed to its outstanding performance. Finally, the OCI-LSTM model stands out because it addresses overfitting, optimises feature selection, and incorporates the GA for increased temporal modelling. The results of the experiments and the many comparisons demonstrate its effectiveness in predicting cardiovascular disease.

Powerful characteristics accuracy of the OCI-LSTM is enhanced by subset choice, network design problem resolution, iterative model improvement utilising optimisation approaches, and extensive comparison with other models. The model has significant potential for practical use given its outstanding performance even with little data.

The OCI-LSTM consistently outperforms the DNN and DBN with regard to of precision, tolerance, F1 index, and accuracy, proving that it is better in CVD prediction. Because of its improved specificity and sensitivity, which show that it is effective in lowering incorrect positives and negatives, the OCI-LSTM is a dependable model for CVD prediction. By exhibiting an excellent balance between accuracy & recall, the high F1 score demonstrates the OCI-LSTM's capacity to capture genuine positive cases while reducing erroneous positives and inaccurate negative outcomes. Based on these metrics, the OCI-LSTM appears to be an effective framework of CVD estimations with a high level of accuracy and reliability overall.

The OCI-LSTM improves CVD prediction accuracy by meticulous feature selection and appropriate network architecture, providing clinicians with a reliable tool. By precisely identifying positive and negative cases, the model's high specificity and sensitivity significantly lower the probability of diagnostic errors. Its capacity to maintain a moderate F1 rating exemplifies its accuracy and memory by allowing practitioners to detect genuine positive circumstances with a minimum of incorrect negatives and positives. The OCI LSTM's efficiency increases its practical use, providing physicians with a trustworthy resource for anticipating when to step in and how to personalise therapy for each individual patient.

CONCLUSION

An OCI-LSTM model efficiently manages underfitting and overfitting, two issues involved with network generalisation, by combining choosing attributes with the SSA and improving the network design with the GA. The OCI-LSTM model repeatedly outperforms the DNN and DBN systems in terms of accuracy when

tested with both the whole feature set and the optimised subset. Analytical and evaluation studies demonstrate the OCI-LSTM's significance in contrast to its rivals. With a staggering accuracy of 98.12%, the OCI-LSTM shows incredible promise in supporting physicians in making informed decisions on the prognosis of heart disease. Verifying the significance of the chosen subset of parameters for cardiovascular health also involves interaction with field experts. Tuning hyperparameters such as learning rates & dropout rates might make the OCI-LSTM more resilient and generalisable to other datasets. An intuitive layout that integrates with existing healthcare systems might facilitate its usage by medical personnel, hence expanding its usefulness in hospitals and clinics.

REFERENCES

- Jenefer, G., A. J. Deepa, and M. Mary Linda. "Diabetic prediction and classification of risk level using ODDTADC method in big data analytics." *Journal of Combinatorial Optimization* 47.5 (2024): 1-31.
- Hussien, Abdelazim G., et al. "An in-depth survey of the artificial gorilla troops optimizer: outcomes, variations, and applications." *Artificial Intelligence Review* 57.9 (2024): 246.
- Huang, Tangsen, Xiangdong Yin, and Ensong Jiang. "Exploring optimization strategies for support vector regression networks in predicting power consumption." *Electrical Engineering* (2024): 1-17.
- Bhoi, Akash Kumar, et al., eds. *Cognitive and Soft Computing Techniques for the Analysis of Healthcare Data*. Academic Press, 2022.
- Siddiqui, Shah Sufi Nesar Uddin. *An Integrated Approach to A Next-Generation Telemedicine and Health Advice System (ANGTHAS) with Artificial Intelligence*. Diss. University of Portsmouth, 2023.
- Datta, Shoumen. "AI (Artificial Intelligence is Incorrect)." (2024).
- da Costa, D. "6G: Vision, Applications, and Challenges Daniel Benevides da Costa, Qiyang Zhao, Marwa Chafii, Faouzi Bader, and Mérouane Debbah." *Fundamentals of 6G Communications and Networking* (2023): 15.
- Kumar, T., and Raja Santhosh. "Design and Dynamic Analysis of Flywheel." *International Conference on Recent Innovations in Civil & Mechanical Engineering, e-ISSN*.
- Habelalmateen, Mohammed I., et al. "Improved Hierarchical Bayesian Optimization Algorithm with LSTM for Crude Oil Price Forecasting." *2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*. IEEE, 2024.
- Jain, Akanksha, and S. C. Gupta. "Optimal placement of distributed generation in the power distribution system and evaluating the losses and voltage using machine learning algorithms." *Frontiers in Energy Research* 12 (2024): 1378242.
- Jayan, Bijili, Tamilarasi Ganesan, and Binu Bhaskara Kurup. "Enhancing IoT network security through advanced data preprocessing and hybrid firefly-salp swarm optimized deep CNN-based intrusion detection." *ITEGAM-JETIA* 10.47 (2024): 73-82.
- Hussien, Abdelazim G., et al. "An in-depth survey of the artificial gorilla troops optimizer: outcomes, variations, and applications." *Artificial Intelligence Review* 57.9 (2024): 246.
- BramahHazela, et al. "Machine Learning: Supervised Algorithms to Determine the Defect in High-Precision Foundry Operation." *Journal of Nanomaterials* 2022.1 (2022): 1732441.
- Gaikar Vilas Bhau, Radhika Gautamkumar Deshmukh, T. Rajasanthosh kumar, Subhadip Chowdhury, Y. Sesharao, Yermek Abilmazhinov, IoT based solar energy monitoring system, Materials Today: Proceedings, Volume 80, Part 3, 2023, Pages 3697-3701, ISSN 2214-7853,
- https://doi.org/10.1016/j.matpr.2021.07.364Guo, Guoqing, Peng Liu, and Yuchen Zheng. "Early energy performance analysis of smart buildings by consolidated artificial neural network paradigms." *Heliyon* 10.4 (2024).
- Jenefer, G., A. J. Deepa, and M. Mary Linda. "Diabetic prediction and classification of risk level using ODDTADC method in big data analytics." *Journal of Combinatorial Optimization* 47.5 (2024): 1-31.
- Bristi, Swarnali Deb, et al. "A Meta-Heuristic Sustainable Intelligent Internet of Things Framework for Bearing Fault Diagnosis of Electric Motor under Variable Load Conditions." *Sustainability* 15.24 (2023): 16722.
- Marzbani, Fatemeh, and Akmal Abdelfatah. "Economic dispatch optimization strategies and problem formulation: A comprehensive review." *Energies* 17.3 (2024): 550.
- Mostafaei, Hasan. "Modal Identification Techniques for Concrete Dams: A Comprehensive Review and Application." *Sci* 6.3 (2024).
- Berghout, Tarek, et al. "A Neural Network Weights Initialization Approach for Diagnosing Real Aircraft Engine Inter-Shaft Bearing Faults." *Machines* 11.12 (2023): 1089.