

Manta Ray Foraging Optimization with Machine Learning Based Biomedical Data Classification

Amal Al-Rasheed¹, Jaber S. Alzahrani², Majdy M. Eltahir³, Abdullah Mohamed⁴,
Anwer Mustafa Hilal^{5,*}, Abdelwahed Motwakel⁵, Abu Sarwar Zamani⁵ and Mohamed I. Eldesouki⁶

¹Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, 11671, Saudi Arabia

²Department of Industrial Engineering, College of Engineering at Alqunfudah, Umm Al-Qura University, Mecca, 24382, Saudi Arabia

³Department of Information Systems, College of Science & Art at Mahayil, King Khalid University, Abha, 62529, Saudi Arabia

⁴Research Centre, Future University in Egypt, New Cairo, 11845, Egypt

⁵Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, AlKharj, Saudi Arabia

⁶Department of Information System, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, AlKharj, Saudi Arabia

*Corresponding Author: Anwer Mustafa Hilal. Email: a.hilal@psau.edu.sa

Received: 12 March 2022; Accepted: 19 April 2022

Abstract: The biomedical data classification process has received significant attention in recent times due to a massive increase in the generation of health-care data from various sources. The developments of artificial intelligence (AI) and machine learning (ML) models assist in the effectual design of medical data classification models. Therefore, this article concentrates on the development of optimal Stacked Long Short Term Memory Sequence-to-Sequence Autoencoder (OSAE-LSTM) model for biomedical data classification. The presented OSAE-LSTM model intends to classify the biomedical data for the existence of diseases. Primarily, the OSAE-LSTM model involves min-max normalization based pre-processing to scale the data into uniform format. Followed by, the SAE-LSTM model is utilized for the detection and classification of diseases in biomedical data. At last, manta ray foraging optimization (MRFO) algorithm has been employed for hyperparameter optimization process. The utilization of MRFO algorithm assists in optimal selection of hyperparameters involved in the SAE-LSTM model. The simulation analysis of the OSAE-LSTM model has been tested using a set of benchmark medical datasets and the results reported the improvements of the OSAE-LSTM model over the other approaches under several dimensions.

Keywords: Biomedical data classification; deep learning; manta ray foraging optimization; healthcare; machine learning; artificial intelligence



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

Recently, research in computer aided intelligent systems for healthcare domains becomes an interesting and essential process [1]. Generally, the physicians utilize their knowledge depending upon patient indications and the long-established diagnoses. As such, prognostic importance of indications towards specific illnesses and symptomatic exactness of a patient are exceptionally subject to a doctor's insight. Since medical information and therapy treatment progressed quickly, for example, the event of new sicknesses and the accessibility of new medications, it is tedious for a doctor to stay updated with all new information and improvement in clinical practices [2]. Due to the developments of computing techniques, it becomes easier to attain and save massive quantities of digital data. Thusly, the organization of modernized medical decision support system turns into a suitable way to deal with helping doctors to quickly and precisely analyze patients [3,4]. In any case, various issues must be defeated before designing a decision support system within the sight of vulnerability and imprecision.

While medical specialists' information and experience are significant, going from evaluating a patient's condition to making a determination, progresses in artificial intelligence (AI) strategies have opened up the way for medical experts to take advantage of electronic shrewd frameworks for decision making in their working environment [5]. Computerized intelligent systems find meaningful in helping the doctor to show up at an educated choice rapidly, for example by gaining from comparable previous cases in an enormous database of electronic patient records and gathering the conclusion for the current patient with appropriate avocations [6]. The benefits of utilizing such smart frameworks incorporate expanding determination precision and, simultaneously, decreasing time and expenses related to patient treatment.

AI models have been created to help different medical dynamic errands. There is an expanding interest in intelligent design of AI and data mining procedures for helping in biomedical examinations as well as in clinical decision making [7]. Conventionally, statistical learning approaches are designed to be performed on the data of previous cases in the diagnosis models which can be utilized for forthcoming cases [8]. They may be utilized to help doctors in directing their choices, and are now and then displayed to beat the specialists' forecast exactness. Besides, such models can find already unnoticed relations between the factors and result in further developing information and comprehension of the condition. Such revelations might bring about better medicines or preventive techniques. Considering that prescient models figure forecasts in view of data of a specific patient, they are additionally encouraging instruments for accomplishing the objective of customized medication [9,10].

This article concentrates on the development of optimal Stacked Long Short Term Memory Sequence-to-Sequence Autoencoder (OSAE-LSTM) model for biomedical data classification. The presented OSAE-LSTM model majorly focuses on the classification of biomedical data for the existence of diseases. Primarily, the OSAE-LSTM model involves min-max normalization based pre-processing to scale the data into uniform format. Followed by, the SAE-LSTM model is utilized for the detection and classification of diseases in biomedical data. At last, the manta ray foraging optimization (MRFO) algorithm has been employed for hyperparameter optimization process. The simulation analysis of the OSAE-LSTM model has been tested using a set of benchmark medical datasets.

2 Literature Review

In [11], a hybridization model has been developed by the use of simulated annealing (SA) and Rao algorithm (RA) to select optimum subset of genes and cancer classification. The SA operates

like a local searching mechanism and RA functions as a globalized model. An optimum gene subset and categorizing cancer. The SA works as a local searching approach and RA work as a global optimized structure. The reason to integrate SA from RA is to enhance the exploitation ability of RA. The presented approach has 2 phases. During the primary phase, minimum redundancy maximum relevance (mRMR) was utilized for choosing the relevant gene subsets in the microarray data set. Afterward, the SA is hybridized with RA for improving the quality of solutions then all iterations of RA.

The authors in [12] examine a novel adaptive technique named Feature Selection(FS)-Seven Spot Ladybird Optimization Algorithm (FS-SLOA) that is a metaheuristic FS technique dependent upon the foraging behavior of 7 spot ladybird. A novel effectual approach was executed for determining an optimum subset feature that attains the maximal accuracy classifier utilizing 3 classifications. In [13], a novel hybrid classification termed fruit-fly optimization (FFO) based extreme learning machine (ELM) was presented for classifying the biomedical data. The projected classification efficiency is also related to different classifications namely support vector machine (SVM) and ELM. These classifications were validated utilizing different performance indices.

In [14], a novel wrapper FS method was presented dependent upon the chimp optimization algorithm (ChOA) for bio-medical data classifier. The ChOA is a recently presented meta-heuristic approach whose ability to resolve FS problems is not examined yet. In 2 binary variations of ChoA are presented to FS problem. During the initial technique, 2 transfer functions (S-shaped and V-shaped) were employed for converting the continuous version of ChoA to binary.

The authors in [15] propose a novel random vector functional link with ε -insensitive Huber loss function (ε -HRVFL) for biomedical data classifier problem. The optimized issue of ε -HRVFL was reworked as strongly convex minimized problems with an easy function iterative method for determining solution. For having an optimum deal of scope of the biomedical data classifier problem and potential solution, it is shown experimentally with 3 distinct kinds of label noise from biomedical dataset and some non-biomedical data sets.

3 The Proposed Model

In this article, a novel OSAE-LSTM model has been developed for the effectual classification of biomedical data for the existence of diseases. The proposed OSAE-LSTM model encompasses a series of operations such as min-max normalization, SAE-LSTM classification, and MRFO based hyperparameter tuning. The utilization of MRFO algorithm assists in optimal selection of hypermeters involved in the SAE-LSTM model. Fig. 1 illustrates the block diagram of OSAE-LSTM technique.

3.1 Min-Max Normalization

At the preliminary stage, the OSAE-LSTM model undergoes min-max normalization based pre-processing to scale the data into uniform format. The min-max normalization process undergoes mathematical formulation by the use of Eq. (1):

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

x' represents normalizing value, x implies the new instance value, x_{max} denotes maximum samples, $\tau_2 = \tau_4 = 1\mu s$ suggests minimum samples. The inverse normalization of the presented method is given in Eq. (2):

$$x = x_{min} + (x_{max} - x_{min}) x' \quad (2)$$

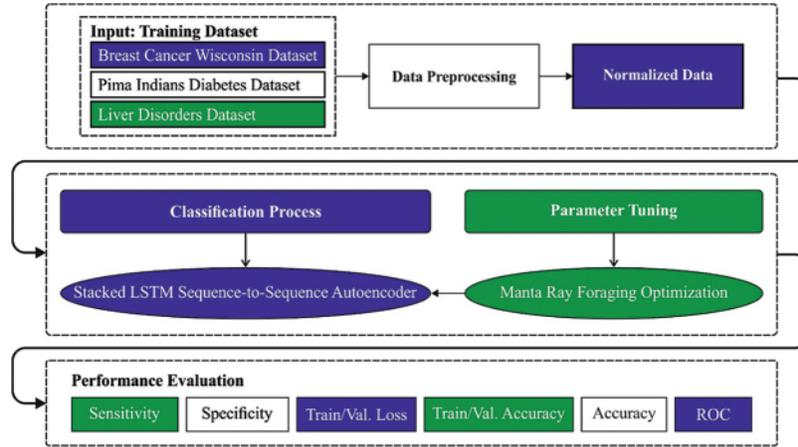


Figure 1: Block diagram of OSAE-LSTM technique

3.2 Process Involved in SAE-LSTM Based Classification

In this study, the SAE-LSTM model is utilized for the detection and classification of diseases in biomedical data. The SAE-LSTM model works as a prototype of a sequence-to-sequence (*seq2seq*) method. The *Seq2seq* paradigm is newly developed commonly in the domain of machine translation and is composed of 2 parts as encoding and decoding [16,17]. The data are established by the encoding that compresses it as to single vector. The vector at present was recognized as context vector (CV), and decoding utilizes it for creating a resultant sequence. The recurrent neural network (RNN) or LSTM was utilized by the encoded for transforming input as to hidden state vector. The encoder resultant vector is state-of-the-art RNN cell hidden state. The encoded send the CV to decoder. The encoding CV was employed as the decoding network initial hidden state, and resultant value of preceding time step is sent to the next LSTM unit as input to progressive forecast.

The mathematical an encoded \emptyset has been created by input as well as hidden layers that compress input data x in a high dimension representation as to low dimension representation Z . Meanwhile, a decoded Ψ was made by the hidden as well as output layers that regenerates the input data x' in the suitable codes. This alteration in the *seq2seq* learning is represented mathematically by the typical neural network (NN) function passed with sigmoid activation function σ (Eq. (3)).

$$\emptyset: X \rightarrow Z \quad (3)$$

$$\chi \mapsto \varphi(x) = \sigma(Wx + b) : = z$$

$$\Psi: Z \rightarrow Z \quad (4)$$

$$z \mapsto \Psi(x) = \sigma(\tilde{W}_z + \tilde{b}) : = \chi'$$

whereas W refers the weighted matrices and b denotes the bias [17].

The encoded and decoded networks of LSTM *seq2seq* method employed for prediction. To utilize this *seq2seq* learning from prediction, LSTM layer is stacked on the encoded and decoded parts of models and named as SAELSTM technique. With stacking LSTM, it is capable of improving our methods prediction abilities to understand further complicated representation of our time-series data from the hidden layer with gathering data at several levels. In addition, x and 0 are the input as well as output data, c implies the encoded CV and ht and st represents the hidden state from the encoded and decoded that is corresponding as follows:

$$h_t = LSTM_{enc}(x_t, h_{t-1}) \tag{5}$$

$$h_t = LSTM_{dec}(o_{t-1}, s_{t-1}). \tag{6}$$

All the encoded LSTM layer computes the CV c , and this CV is simulated and sent to all the decoded units.

3.3 Process Involved in MRFO Based Hyperparameter Optimization

At the final stage, the MRFO algorithm has been employed for hyperparameter optimization process which assists in optimal selection of hypermeters involved in the SAE-LSTM model. The MRFO is a bio-inspired new technique which simulates the intelligent foraging performance of manta rays (MRs) and features of its foraging performance. The model was appropriate to our current solar radiation forecast problem provided that MRs on that the MRFO is created, have 3 various foraging approaches which are utilized for searching for food that procedure the vital search methods of MRFO for optimizing the solution of our presented solar radiation forecast problem [18]. The mathematical process of chain foraging was signifying as:

$$M_m^* = \begin{cases} M_m + (M_B - M_m)(r + \sigma) & \text{if } m = 1 \\ M_m + r(M_{m-1} - M_m) + \sigma(M_B - M_m) & \text{if } m \neq 1 \end{cases} \tag{7}$$

$$\sigma = 2r\sqrt{|\log(r)|} \tag{8}$$

In which (M_m) stands for the individual MR (m), r refers the arbitrary uniformly distributed number from the range of zero and one. M^* and MB defines the novel or optimum position of MR from the population, σ denotes the weighted co-efficient as function of all the iterations. Fig. 2 depicts the flowchart of MRFO technique.

It can be apparent in Eq. (8) that the preceding MR from the chain and spatial place of strongest plankton obviously determine the position upgrade method from the chain foraging. Cyclone foraging was separated as to 2 parts. The 1st half concentrations on improving the exploration and is upgraded as:

$$M_m^* = \begin{cases} M_R + (M_R - M_m)(r + \beta) & \text{if } m = 1 \\ M_m + r_1(M_{m-1} - M_m) + \beta(M_R - M_m) & \text{if } m \neq 1 \end{cases} \tag{9}$$

whereas M_R signifies the individual generated arbitrarily:

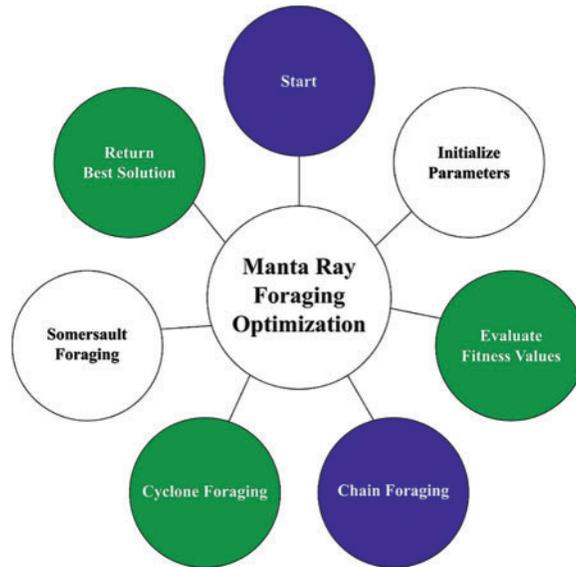


Figure 2: Flowchart of MRFO technique

$$M_R = M^{\min} + r_1 (M^{\max} - M^{\min}). \quad (10)$$

The adaptive weighted co-efficient (β) was diverse as:

$$\beta = 2e^{r_2 \frac{Iter_m - Iter_m + 1}{Iter_m}} \sin(2\pi r_2) \quad (11)$$

In which $Iter$ implies the present iteration and arbitrary uniformly distributed number, and r_2 is over of zero and one. The 2nd half concentrate on enhancing the exploitation, thus the upgrade is as per:

$$M_m^* = \begin{cases} M_B + (M_B - M_m)(r_1 + \beta) & \text{if } m = 1 \\ M_B + r_1(M_{m-1} - M_m) + \beta(M_B - M_m) & \text{if } m \neq 1 \end{cases} \quad (12)$$

Somersault foraging: The ending foraging approach with MRs determining the food supply and exploiting backward somersaults for circling the plankton for attracting. Somersaulting is local, spontaneous, cyclical, and periodic act which MRs utilize for maximizing their food intake. The 3rd approach is where an upgrade of all individuals takes place around an optimum position:

$$M_m^* = M_m + S(r_3 M_B - r_4 M_m). \quad (13)$$

In Eq. (13), S represents the somersault co-efficient ($S = 2$) adjusting the domain of MRs, r_3 and r_4 are arbitrary numbers in the range of zero and one. According to an arbitrarily created number, the MRFO technique is switched amongst chain as well as cyclone foragings. Afterward, the summersault foraging gets act for updating individual's present positions utilizing an optimum solution obtainable at the time. These 3 various foraging procedures are utilized interchangeably for achieving the global optimal solution of optimized problem, so sufficient the already decided end condition.

The MRFO method made a FF for reaching higher classifier performance. It defines a positive integer for demonstrating the best result of candidate solutions. Under this work, the minimized

classification error rate is regarded as FF is given in Eq. (14). The best result is a less error rate and worst outcome reaches a higher error rate.

$$\begin{aligned}
 fitness(x_i) &= ClassifierErrorRate(x_i) \\
 &= \frac{numberofmisclassifiedsamples}{Totalnumberofsamples} * 100
 \end{aligned}
 \tag{14}$$

4 Results and Discussion

The performance validation of the OSAE-LSTM model is performed using three benchmark datasets [19]. Firstly, the Breast Cancer Wisconsin (BCW) dataset with 569 samples with 32 features is used. Next, the PIMA Indians dataset (PID) includes 768 instances with 8 features. Finally, the Liver Disorders dataset comprises 345 samples with 7 attributes.

Tab. 1 offers a detailed classifier outcome of the OSAE-LSTM model on the test BCW dataset. The experimental values highlighted that the OSAE-LSTM model has accomplished maximum classifier results. For instance, with run-1, the OSAE-LSTM model has offered $accu_y$, $sens_y$, $spec_y$, and receiver operating characteristic (ROC) of 97.07%, 98.27%, 98.49%, and 97.59% respectively. At the same time, run-2, the OSAE-LSTM technique has offered $accu_y$, $sens_y$, $spec_y$, and ROC of 97.16%, 97.77%, 98.13%, and 98.48% correspondingly. Moreover, run-4, the OSAE-LSTM model has offered $accu_y$, $sens_y$, $spec_y$, and ROC of 97.49%, 98.61%, 97.99%, and 99% respectively. Likewise, run-5, the OSAE-LSTM approach has offered $accu_y$, $sens_y$, $spec_y$, and ROC of 98.61%, 97.13%, 97.36%, and 98.15% correspondingly.

Table 1: Result analysis of OSAE-LSTM technique with distinct measures on BCW dataset

Breast cancer wisconsin dataset				
No. of runs	Accuracy	Sensitivity	Specificity	ROC
Run-1	97.07	98.27	98.49	97.59
Run-2	97.16	97.77	98.13	98.48
Run-3	97.95	97.00	98.20	97.31
Run-4	97.49	98.61	97.99	99.00
Run-5	98.61	97.13	97.36	98.15
Average	97.80	97.63	97.92	98.24

Fig. 3 illustrates the training and validation accuracy inspection of the OSAE-LSTM technique on BCW dataset. The figure conveyed that the OSAE-LSTM model has offered maximum training/validation accuracy on classification process.

Next, Fig. 4 exemplifies the training and validation loss inspection of the OSAE-LSTM approach on BCW dataset. The figure exposed that the OSAE-LSTM model has offered decreased training/accuracy loss on the classification process of test data.

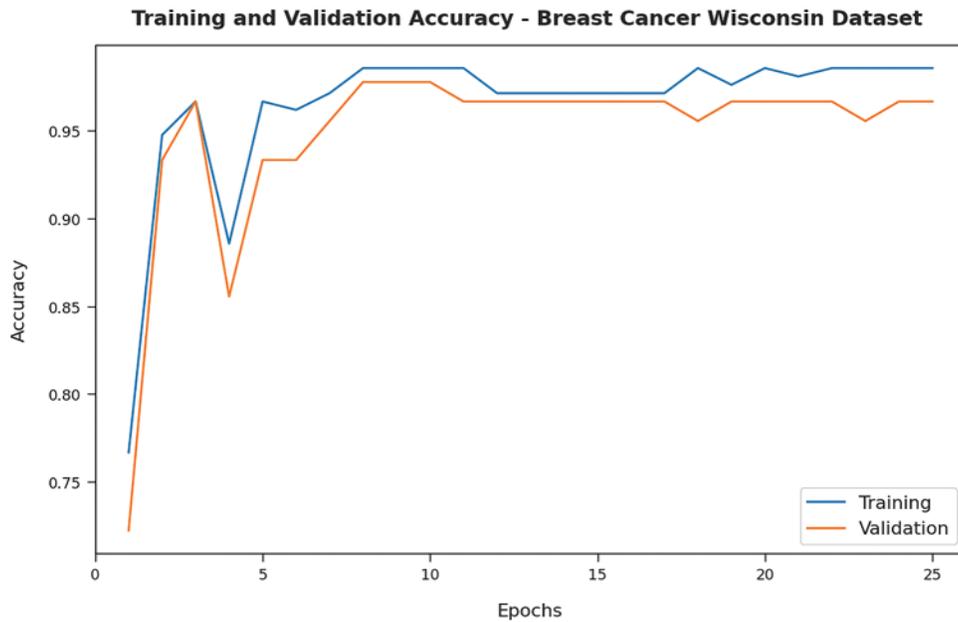


Figure 3: Accuracy analysis of OSAE-LSTM technique on BCW dataset

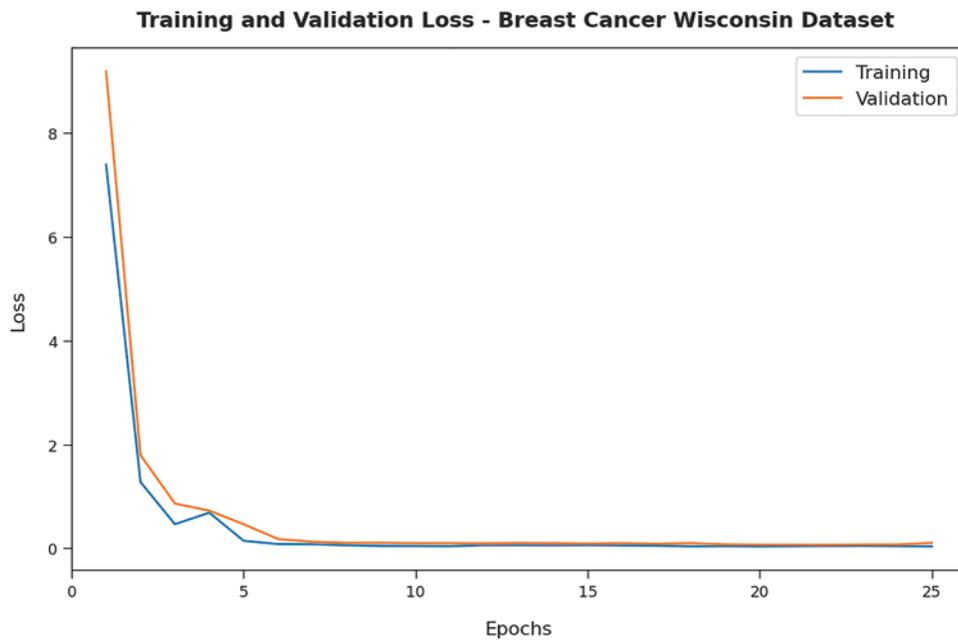


Figure 4: Loss analysis of OSAE-LSTM technique on BCW dataset

Tab. 2 and Fig. 5 demonstrate the comparison study of the OSAE-LSTM model with recent models such as fuzzy min-max (FMM), FMM with Classification And Regression Tree (CART), FMM-CART-random forest (RF) on BCW dataset. The table values indicated that the FMM-CART model has resulted to $sens_y$, $spec_y$, $accu_y$, and ROC of 95.01%, 90.87%, 84.82%, and 97.29% respectively. Along with that, the FMM model has obtained slightly increased outcome with $sens_y$, $spec_y$, $accu_y$, and

ROC of 94.87%, 94.99%, 94.27%, and 96.57% respectively. Though the FMM-CART-RF model has resulted to $sens_y$, $spec_y$, $accu_y$, and ROC of 97.41%, 97.12%, 97.26%, and 98.10%, the OSAE-LSTM model has reached maximum performance with $sens_y$, $spec_y$, $accu_y$, and ROC of 97.63%, 97.92%, 97.80%, and 98.24% respectively.

Table 2: Comparative analysis of OSAE-LSTM technique with existing algorithm on BCW dataset

Methods	Sensitivity	Specificity	Accuracy	ROC
FMM	94.87	94.99	94.27	96.57
FMM-CART	95.01	90.87	84.82	97.29
FMM-CART-RF	97.41	97.12	97.26	98.01
OSAE-LSTM	97.63	97.92	97.80	98.24

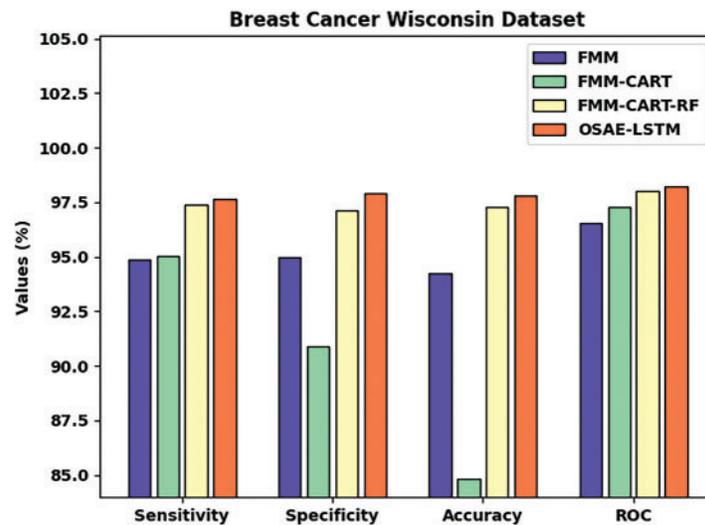


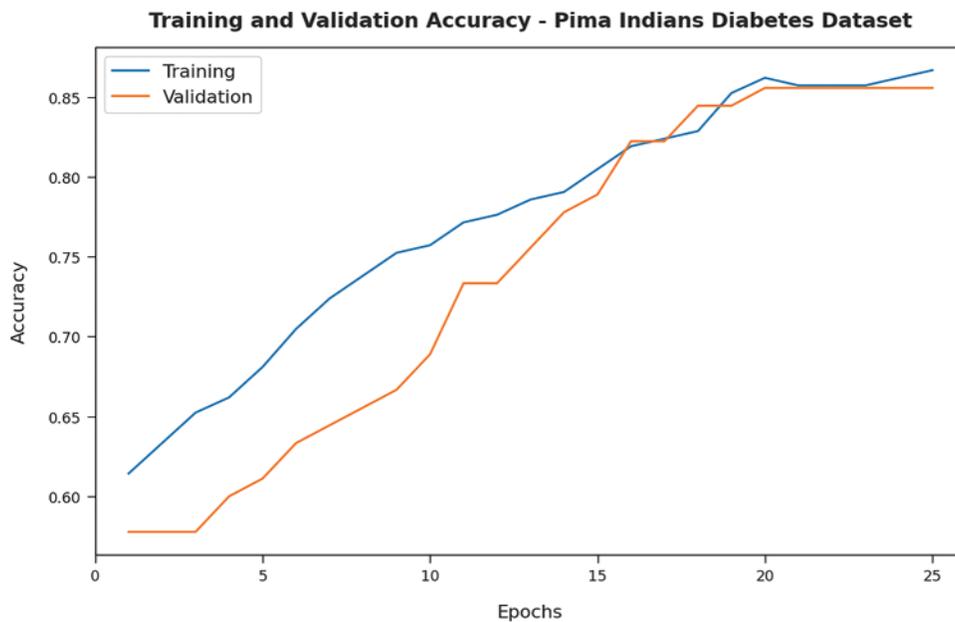
Figure 5: Comparative analysis of OSAE-LSTM technique on BCW dataset

Tab. 3 provides a detailed classifier outcome of the OSAE-LSTM technique on the test PID dataset. The experimental values highlighted that the OSAE-LSTM algorithm has accomplished maximum classifier results. For instance, with run-1, the OSAE-LSTM methodology has accessible $accu_y$, $sens_y$, $spec_y$, and ROC of 82.98%, 80.54%, 80.77%, and 76.66% correspondingly. Concurrently, run-2, the OSAE-LSTM approach has offered $accu_y$, $sens_y$, $spec_y$, and ROC of 81.27%, 79.53%, 82.57%, and 76.13% correspondingly. Additionally, run-4, the OSAE-LSTM model has offered $accu_y$, $sens_y$, $spec_y$, and ROC of 82.41%, 82.45%, 80.37%, and 76.46% individually. Likewise, run-5, the OSAE-LSTM technique has obtainable $accu_y$, $sens_y$, $spec_y$, and ROC of 82.33%, 80.91%, 81.51%, and 76.87% correspondingly.

Table 3: Result analysis of OSAE-LSTM technique with distinct measures on PID dataset

Pima indians diabetes dataset				
No. of runs	Accuracy	Sensitivity	Specificity	ROC
Run-1	82.98	80.54	80.77	76.66
Run-2	81.27	79.53	82.57	76.13
Run-3	82.66	81.12	82.34	78.22
Run-4	82.41	82.45	80.37	76.46
Run-5	82.33	80.91	81.51	76.87
Average	82.98	80.54	80.77	76.66

Fig. 6 showcases the training and validation accuracy inspection of the OSAE-LSTM technique on PID dataset. The figure conveyed that the OSAE-LSTM system has offered maximum training/validation accuracy on classification process.

**Figure 6:** Accuracy analysis of OSAE-LSTM technique on PID dataset

Afterward, Fig. 7 represents the training and validation loss inspection of the OSAE-LSTM approach on PID dataset. The figure showing that the OSAE-LSTM technique has offered decreased training/accuracy loss on the classification process of test data.

Tab. 4 and Fig. 8 showcase the comparison study of the OSAE-LSTM technique with recent techniques on PID dataset. The table values referred that the FMM-CART model has resulted to $sens_y$, $spec_y$, $accu_y$, and ROC of 76.21%, 60.16%, 70.87%, and 68.51% correspondingly. Similarly, the FMM algorithm has obtained somewhat improved outcome with $sens_y$, $spec_y$, $accu_y$, and ROC of 73.24%, 57.68%, 69.57%, and 66.46% respectively. Followed by, the FMM-CART-RF system has resulted to

$sens_y$, $spec_y$, $accu_y$, and ROC of 79.54%, 66.39%, 78.59%, and 73.34%, the OSAE-LSTM approach has reached superior performance with $sens_y$, $spec_y$, $accu_y$, and ROC of 80.91%, 81.51%, 82.33%, and 76.87% respectively.

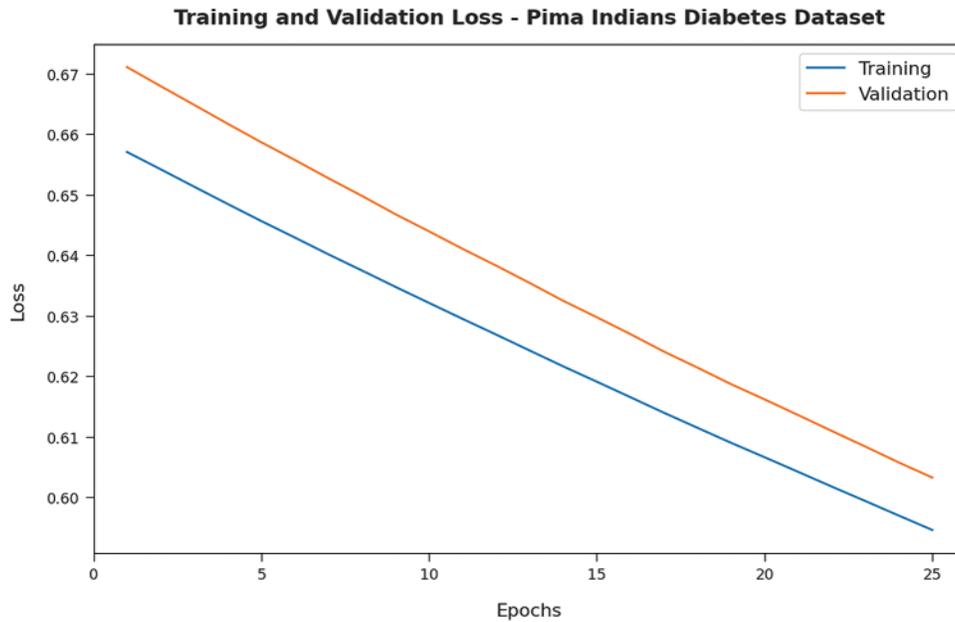


Figure 7: Loss analysis of OSAE-LSTM technique on PID dataset

Table 4: Comparative analysis of OSAE-LSTM technique with existing algorithm on PID dataset

Methods	Sensitivity	Specificity	Accuracy	ROC
FMM	73.24	57.68	69.57	66.46
FMM-CART	76.21	60.16	70.87	68.51
FMM-CART-RF	79.54	66.39	78.59	73.34
OSAE-LSTM	80.91	81.51	82.33	76.87

Tab. 5 provides a detailed classifier outcome of the OSAE-LSTM model on the test liver disorders dataset. The experimental values highlighted that the OSAE-LSTM system has accomplished maximal classifier results. For instance, with run-1, the OSAE-LSTM approach has accessible $accu_y$, $sens_y$, $spec_y$, and ROC of 90.79%, 92.41%, 90.07%, and 90.85% correspondingly. Simultaneously, run-2, the OSAE-LSTM model has offered $accu_y$, $sens_y$, $spec_y$, and ROC of 90.55%, 90.90%, 90.07%, and 92% correspondingly. Besides, run-4, the OSAE-LSTM technique has offered $accu_y$, $sens_y$, $spec_y$, and ROC of 91.96%, 91.75%, 90.77%, and 90.56% respectively. Also, run-5, the OSAE-LSTM method has obtainable $accu_y$, $sens_y$, $spec_y$, and ROC of 92.33%, 91.69%, 92.17%, and 92.85% respectively.

Fig. 9 illustrates the training and validation accuracy inspection of the OSAE-LSTM technique on liver disorders dataset. The figure conveyed that the OSAE-LSTM method has offered maximum training/validation accuracy on classification process.

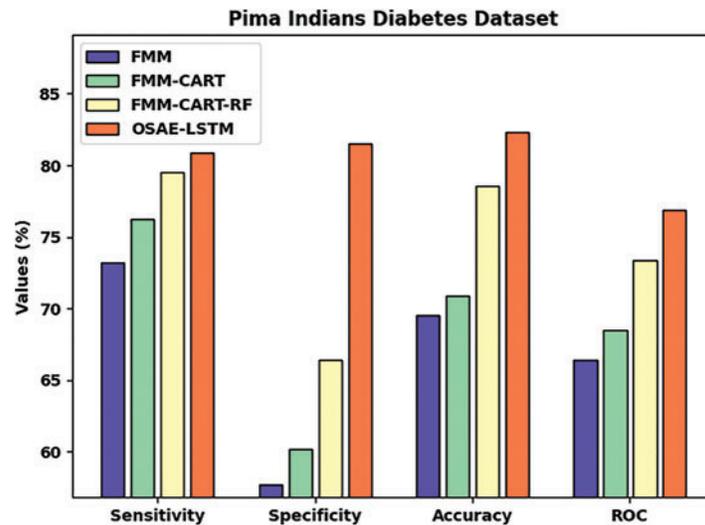


Figure 8: Comparative analysis of OSAE-LSTM technique on PID dataset

Table 5: Result analysis of OSAE-LSTM technique with distinct measures on liver disorders dataset

Liver disorders dataset				
No. of runs	Accuracy	Sensitivity	Specificity	ROC
Run-1	90.79	92.41	90.07	90.85
Run-2	90.55	90.90	90.07	92.00
Run-3	92.64	90.46	90.35	92.51
Run-4	91.96	91.75	90.77	90.56
Run-5	92.33	91.69	92.17	92.85
Average	91.87	91.20	90.84	91.98

Then, [Fig. 10](#) demonstrates the training and validation loss inspection of the OSAE-LSTM approach on liver disorders dataset. The figure revealed that the OSAE-LSTM model has offered decreased training/accuracy loss on the classification process of test data.

[Tab. 6](#) and [Fig. 11](#) illustrate the comparison study of the OSAE-LSTM method with recent models on liver disorders dataset [20]. The table values indicated that the FMM-CART model has resulted to $sens_y$, $spec_y$, $accu_y$, and ROC of 84.87%, 82.17%, 84.28%, and 87.91% respectively. Besides, the FMM system has obtained slightly increased outcome with $sens_y$, $spec_y$, $accu_y$, and ROC of 63.38%, 70.69%, 64.27%, and 71.10% correspondingly. Moreover, the FMM-CART-RF model has resulted to $sens_y$, $spec_y$, $accu_y$, and ROC of 88.77%, 88.09%, 88.16%, and 90.84%, the OSAE-LSTM approach has reached maximal performance with $sens_y$, $spec_y$, $accu_y$, and ROC of 91.20%, 90.84%, 91.87%, and 91.98% correspondingly.

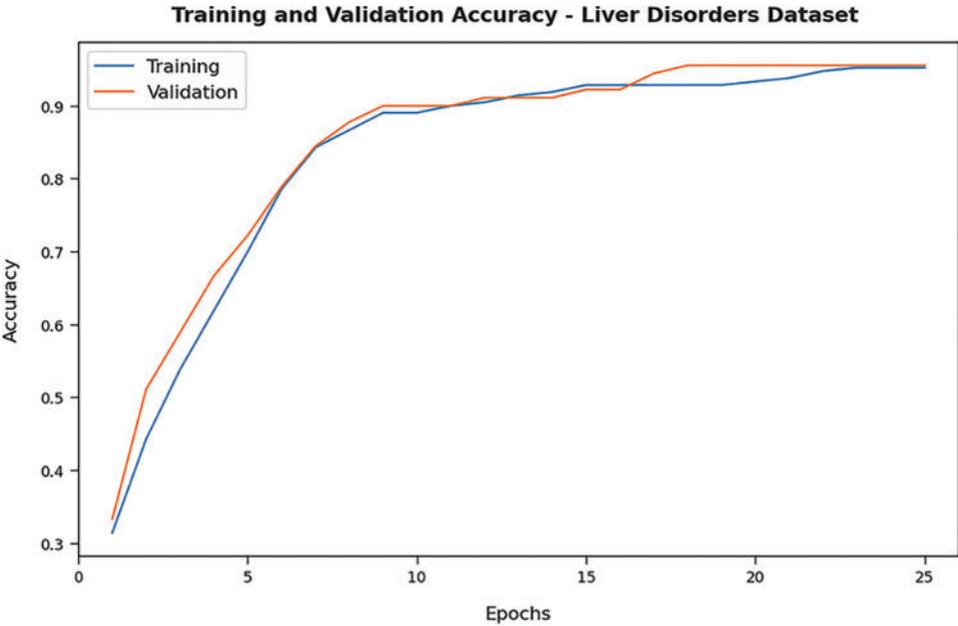


Figure 9: Accuracy analysis of OSAE-LSTM technique on liver disorders dataset

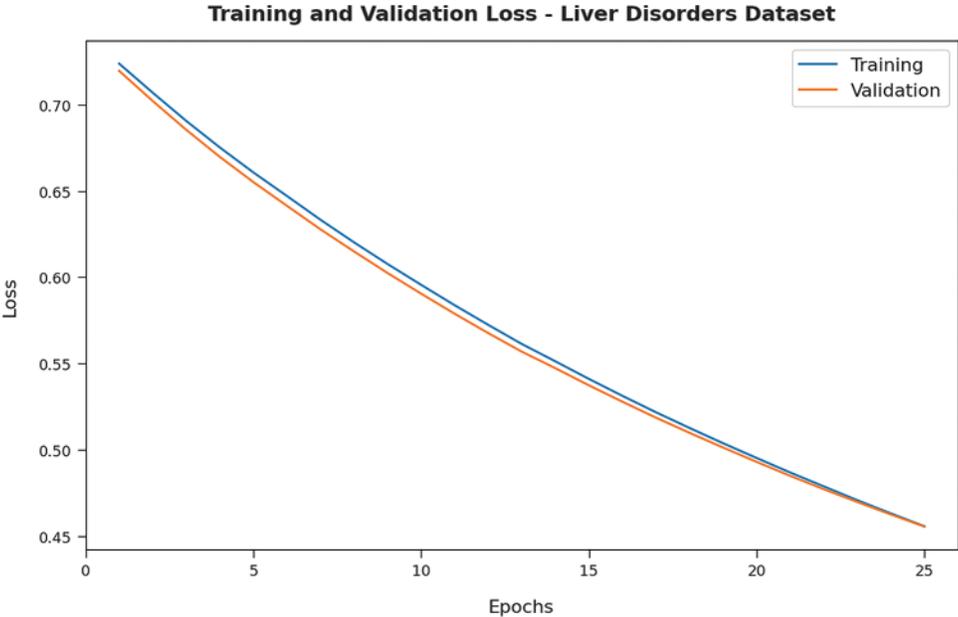
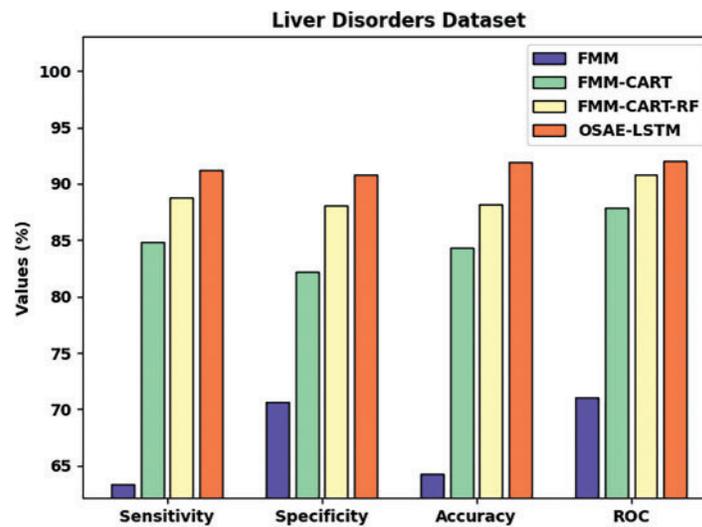


Figure 10: Loss analysis of OSAE-LSTM technique on liver disorders dataset

Table 6: Comparative analysis of OSAE-LSTM technique with existing algorithm on liver disorders dataset

Methods	Sensitivity	Specificity	Accuracy	ROC
FMM	63.38	70.69	64.27	71.10
FMM-CART	84.87	82.17	84.28	87.91
FMM-CART-RF	88.77	88.09	88.16	90.84
OSAE-LSTM	91.20	90.84	91.87	91.98

**Figure 11:** Comparative analysis of OSAE-LSTM technique on liver disorders dataset

5 Conclusion

In this article, a novel OSAE-LSTM model has been developed for the effectual classification of biomedical data for the existence of diseases. The proposed OSAE-LSTM model encompasses a series of operations such as min-max normalization, SAE-LSTM classification, and MRFO based hyperparameter tuning. The utilization of MRFO algorithm assists in optimal selection of hyperparameters involved in the SAE-LSTM model. The simulation analysis of the OSAE-LSTM model has been tested using a set of benchmark medical datasets and the results reported the improvements of the OSAE-LSTM model over the other approaches under several dimensions. Thus, the presented OSAE-LSTM model has been employed for effectual detection and classification of biomedical data. In future, feature selection models can be introduced to reduce the high dimensionality problem that exist in the heterogeneous biomedical data.

Funding Statement: The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work under grant number (RGP 2/158/43). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R235), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors would like to thank the

Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (22UQU4340237DSR06).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] R. S. Bressan, G. Camargo, P. H. Bugatti and P. T. M. Saito, "Exploring active learning based on representativeness and uncertainty for biomedical data classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2238–2244, 2019.
- [2] L. Jena, S. Nayak and R. Swain, Chronic disease risk (CDR) prediction in biomedical data using machine learning approach. In: *Advances in Intelligent Computing and Communication, Lecture Notes in Networks and Systems Book Series*. Vol. 109. Singapore: Springer, pp. 232–239, 2020.
- [3] W. Sun, X. Chen, X. R. Zhang, G. Z. Dai, P. S. Chang *et al.*, "A multi-feature learning model with enhanced local attention for vehicle re-identification," *Computers, Materials & Continua*, vol. 69, no. 3, pp. 3549–3561, 2021.
- [4] W. Sun, L. Dai, X. R. Zhang, P. S. Chang and X. Z. He, "RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring," *Applied Intelligence*, vol. 92, no. 6, pp. 1–16, 2021.
- [5] C. Su, J. Tong, Y. Zhu, P. Cui and F. Wang, "Network embedding in biomedical data science," *Briefings in Bioinformatics*, vol. 21, no. 1, pp. 182–197, 2018.
- [6] E. Momanyi and D. Segera, "A master-slave binary grey wolf optimizer for optimal feature selection in biomedical data classification," *BioMed Research International*, vol. 2021, no. 3, pp. 1–12, 2021.
- [7] F. Deng, J. Huang, X. Yuan, C. Cheng and L. Zhang, "Performance and efficiency of machine learning algorithms for analyzing rectangular biomedical data," *Laboratory Investigation*, vol. 101, no. 4, pp. 430–441, 2021.
- [8] P. Agarwalla and S. Mukhopadhyay, "Comparative analysis of multi-objective algorithms for imbalanced biomedical data classification with tuned classifier," in *2021 5th Int. Conf. on Information Systems and Computer Networks (ISCON)*, Mathura, India, pp. 1–6, 2021.
- [9] H. Luo, M. Li, M. Yang, F. Wu, Y. Li *et al.*, "Biomedical data and computational models for drug repositioning: A comprehensive review," *Briefings in Bioinformatics*, vol. 22, no. 2, pp. 1604–1619, 2020.
- [10] M. T. Islam and L. Xing, "A data-driven dimensionality-reduction algorithm for the exploration of patterns in biomedical data," *Nature Biomedical Engineering*, vol. 5, no. 6, pp. 624–635, 2021.
- [11] S. K. Baliarsingh, K. Muhammad and S. Bakshi, "SARA: A memetic algorithm for high-dimensional biomedical data," *Applied Soft Computing*, vol. 101, no. 4, pp. 107009, 2021.
- [12] N. Bidi and Z. Elberrichi, "Best features selection for biomedical data classification using seven spot ladybird optimization algorithm," *International Journal of Applied Metaheuristic Computing*, vol. 9, no. 3, pp. 75–87, 2018.
- [13] P. Parhi, J. Naik, S. P. Mishra and R. Bisoi, "A hybridized levy flight fruit fly optimization based kernel extreme learning machine for biomedical data classification," in *2020 Int. Conf. on Artificial Intelligence and Signal Processing (AISP)*, Amaravati, India, pp. 1–5, 2020.
- [14] E. Pashaei and E. Pashaei, "An efficient binary chimp optimization algorithm for feature selection in biomedical data classification," *Neural Computing and Applications*, vol. 34, no. 8, pp. 6427–6451, 2022.
- [15] B. B. Hazarika and D. Gupta, "Random vector functional link with ϵ -insensitive Huber loss function for biomedical data classification," *Computer Methods and Programs in Biomedicine*, vol. 215, no. 8, pp. 106622, 2022.
- [16] S. Ghimire, R. C. Deo, H. Wang, M. S. A. Musaylh, D. C. Pérez *et al.*, "Stacked LSTM sequence-to-sequence autoencoder with feature selection for daily solar radiation prediction: a review and new modeling results," *Energies*, vol. 15, no. 3, pp. 1061, 2022.

- [17] K. Cho, D. Bahdanau, D. Bahdanau, F. Bougares, H. Schwenk *et al.*, “Learning phrase representations using rnn encoder-decoder for statistical machine translation,” in *Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, pp. 1724–1734, 2014.
- [18] W. Zhao, Z. Zhang and L. Wang, “Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications,” *Engineering Applications of Artificial Intelligence*, vol. 87, no. 5, pp. 103300, 2020.
- [19] <https://archive.ics.uci.edu/ml/datasets.php>.
- [20] M. Seera and C. P. Lim, “A hybrid intelligent system for medical data classification,” *Expert Systems with Applications*, vol. 41, no. 5, pp. 2239–2249, 2014.