



Connection Weight Optimization in a Long Short-Term Memory Using Whale Optimization Algorithm

¹Abuzairu Ahmad, ²Fatima Umar Zambuk, ³Kabiru Ibrahim Musa, ⁴Mohammed Ajuji, ⁵Useni Datti Emmanuel

^{1&2}Department of Mathematical Sciences, ATBU, Bauchi

³Department of Management Information Technology, ATBU, Bauchi

⁴Department of Computer Science, Gombe State University, Gombe State

⁵Department of Computer Science, School of Science, Federal College of Education Pankshin

ABSTRACT

Many academics have recently developed an interest in the procedure for learning LSTM, as well as recognized as an example trickiest issue in artificial intelligence. In most cases, local optima stagnation and slow convergence are the main drawbacks of the standard training procedure. Therefore, it is possible to rely on the stochastic optimization approach to handle these problems. LSTM training has been done using a variety of evolutionary and swarm-based techniques, but the problem of local minima and MSE still exists. Therefore, it was suggested in this study endeavor to optimize the link weights in a long short-term memory using the whale optimization technique. The study compares and presents the results from several datasets in greater depth. (Bladder, Leukemia, Lung, Ovary, Pancreas, and, separately, Prostate cancer) The classification accuracy, convergence rate, and precision score with the highest performance will be determined using MATLAB R2021a. The recommended WOA trainer is compared to GA. Experimental results show that the proposed WOA came out on top for classification accuracy in four instances (bladder, ovary, leukemia, and prostate cancers), whereas GA came out on top in two instances (lung and pancreas cancers). The proposed model outperforms the other algorithm in terms of average convergence speed (MSE) for the datasets related to bladder and ovary cancer, with average MSEs of 0.01888 and 0.006428, respectively. The GA, on the other hand, came in second place, with average MSEs of 0.006984, 0.007946, 0.009627, and 0.008148, respectively. Due to its extensive exploration and avoidance of local optima, the WOA was able to demonstrate results that were superior to those of the other algorithm in terms of convergence. This study further demonstrates that the suggested trainer can successfully train RNN to classify datasets with varying degrees of difficulty.

ARTICLE INFO

Article History

Received: May, 2025

Received in revised form: June, 2025

Accepted: August, 2025

Published online: September, 2025

KEYWORDS

Training Neural Network, Evolutionary Algorithm, Whale Optimization Algorithm, Long Short-Term Memory, Machine Learning

INTRODUCTION

Models of intelligence and nonparametric mathematics inspired by the biological nervous system are artificial neural networks (ANNs). Artificial neural networks created extensively studied and applied to problems involving classification, pattern

recognition, regression, and forecasting over the last three decades. Aljarah and colleagues (2018) the learning process of ANNs has a large impact on their efficiency. The artificial neural network learning process regarded as one of machine learning's most difficult challenges, and It was only

Corresponding author: Abuzairu Ahmad

✉ abuzairuhammad2020@gmail.com

Department of Mathematical Sciences, ATBU, Bauchi.

© 2025. Faculty of Technology Education. ATBU Bauchi. All rights reserved



recently that drawn attention of large number of researchers.

The most important challenge is the process of training a neural network that it is nonlinear, as well as that the most effective set of primary controlling parameters is unknown (Biases and weights). Stagnation of local optimums and slow convergence speed are the main drawbacks of traditional training algorithms. As a result, the stochastic optimization algorithm is a dependable alternative for mitigating these drawbacks (Aljarah et al, 2018). Deep learning, led by convolutional neural networks (CNN) and recurrent neural networks (RNN), has pushed the boundaries of many computer vision applications (Lipton et al, 2021).

Deep learning (DL) has recently emerged as the fastest-growing trend in big data analysis, having been widely and successfully applied to a variety of computer applications such as sequential data, natural language processing, speech recognition, and image classification Dargan et al, (2020). Because of its superior performance in comparison to traditional learning algorithms. Machine learning techniques are becoming increasingly important as the science paradigm shifts toward data-intensive science. Deep learning, in particular, has proven to be an extremely powerful tool in a variety of fields as a major breakthrough in the field.

REVIEW OF RELATED LITERATURE

Prediction algorithms are classified into approaches to shallow learning and approaches to deep learning. Deep learning has become popular and emerged as a viable option when prediction model performance outperformed computational power. All previously published works on turbotan RUL prediction engines utilizing the same benchmark datasets are compared in order to make a fair and consistent comparison. A brief synopsis of related works' methodology is provided. The first step is to summarize shallow, learning-based approaches. For instance, in Chui et al. (2021), the authors proposed a hybrid discrete Bayesian filter and K-nearest neighbors approach. It had a root-mean-square error (RMSE) average of 27.57. According to the

literature, long- and short-term memory was used in a large proportion of the articles. Long-term prediction using LSTM was lauded by researchers.

To improve the LSTM model, the Adam learning rate optimization algorithm was used (Chui et al., 2021). The configuration of 5 layers with 100 neurons in each layer resulted in maximum efficiency (average RMSE of 19.1). The use of heuristic optimization algorithms undoubtedly results in much faster execution (Edoh, et al., 2020). A single-layer and a multi-layer long-term memory (LSTM) model for weather forecasting with intermediate variables were proposed (Salman et al., 2018). The best LSTM model in this experiment has a number of layers. Pressure variable data is the best intermediate data for LSTM. The validation accuracy can be calculated using the pressure variable, which was 0.8060, and the RMSE was 0.0775.

STATEMENT OF THE PROBLEM

Artificial neural networks (ANNs) with a single hidden layer always have less accuracy than ANNs with multiple hidden layers (Aljarah et al., 2018). Although a wide range of evolutionary and swarm-based algorithms are deployed and investigated in the literature for training MLP, the problem of local minima is still open (Aljarah et al., 2018). Whale optimization algorithm (WOA) utilizes a population of search agents to determine problems. The search process starts with creating a set of random solutions (candidate solutions) for a given problem. It then improves this set until the satisfaction of an end criterion (Mirjalili and Aljarah, 2018).

The main difference between WOA and other algorithms is the rules that improve the candidate solutions in each step of optimization. In fact, WOA mimics the hunting behavior of humpback whales in finding and attacking preys called bubble-net feeding behavior (Aljarah et al., 2018). Motivated by these reasons, on the recent whale optimization algorithm is proposed for training a LSTM with multiple hidden layers. WOA, a novel meta-heuristic algorithm, was first introduced and developed by Mirjalili and Lewis

(2018). WOA is inspired by the bubble-net hunting strategy of humpback whales. In optimization algorithm literature, there is no optimization algorithm that logically proves no-free-lunch (NFL) theorem for solving all optimization problems (Mirjalili et al., 2021). But WOA proved that it can be used for all optimization problems (Mirjalili, 2018).

Aim and Objectives

The aim of this work is to optimize the connection weights in a long short-term memory using whale optimization algorithm. This research has the following objectives to;

1. Propose a novel long short-term memory training technique using whale optimization algorithm.
2. Compare and visualize the results obtained on different datasets in order to find the best classification accuracy, convergent speed, and precision score.
3. Developed WOA-based approach in this research will be evaluated and

tested based on six (6) selected cancer classification datasets.

METHODOLOGY

The primary goal of this research is to maximize the link weights in a long working memory. WOA was employed in the training of an LSTM network with multiple hidden layers, dubbed WOA-LSTM. In WOA-LSTM stores, each search agent is represented as a one-dimensional vector by a potential neural network. A vector is made up of three parts: biases, weights bridging the hidden and output layers, and Weights connecting the hidden and input layers. The length of each vector equals the total number of weights and biases in the network and may be determined using equation 1 where n is the number of input variables and m is the number of neurons in the hidden layer:

$$\text{Individual length} = (n \times m) + (2 \times m) + 1$$

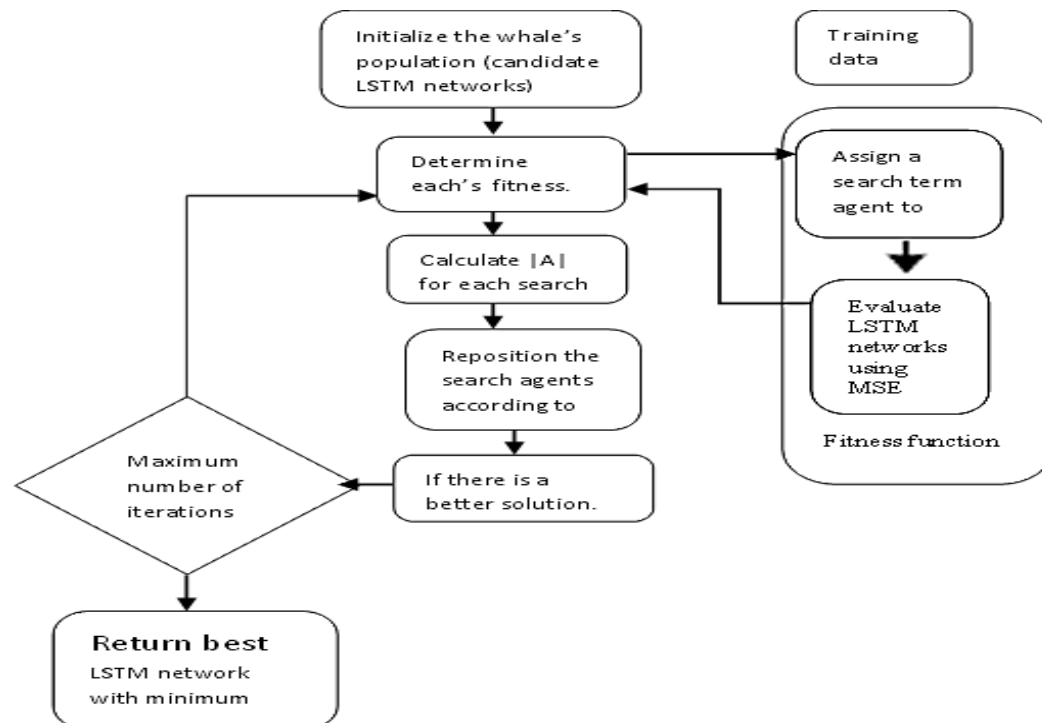


Fig. 1. Frame work of the proposed system

Corresponding author: Abuzairu Ahmad

✉ abuzairuahmad2020@gmail.com

Department of Mathematical Sciences, ATBU, Bauchi.

© 2025. Faculty of Technology Education. ATBU Bauchi. All rights reserved

Figure 3 depicts the WOA-LSTM approach proposed as a foundation for training the LSTM network. WOA is employed in Train an LSTM network with a large number of hidden layers. The WOA-based methods used to train the LSTM network is steps are as follows:

Step 1: Upon initialization, a predefined number of random search agents are created. Every search engine agent may contain an LSTM network.

Step 2: Fitness evaluation entails assessing how well the built LSTM networks perform using a fitness function. Each LSTM network is evaluated after receiving the collection of weights and biases that comprise the axes of the search agents. In this study, the MSE, a fitness function commonly used in evolutionary neural networks, is used. The training algorithm's goal is to identify the LSTM network that has the lowest MSE value calculated from the dataset's training samples.

Step 3: Locations should be updated for the search agents.

Step 4: Repeat steps 2 through 3 for the number of repetitions specified. Step 5: The LSTM The network with the lowest MSE value is then tested on previously unseen data (test or validation samples).

To measure the fitness value of the generated WOA agents, we use the mean square error (MSE) fitness function, which is based on calculating the difference between the actual and predicted values of the generated agents (LSTM) for all training samples. The MSE is given in Equation 2, where y is the actual value, $\hat{y}$ is the predicted value, and n is the number of occurrences in the training dataset.

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

Long Short-Term Memory

LSTM is more accurate than traditional RNNs [Abuzairu]. It was proposed in [Abuzairu]. Memory blocks, as opposed to RNN, are discrete units found in the LSTM recurrent hidden layer [Abuzairu]. Memory blocks are made up of

memory cells with self-connections that record the network temporal state as well as specific multiplicative units called gates that regulate information flow. In the original architecture, each memory block featured three distinct gate types: an input gate, an output gate, and a forget gate [Abuzairu].

Whale Optimization Algorithm

The WOA makes use of a population of search agents tasked with locating the best global solution to optimization problems. Like alternative population-based algorithms, the search process starts with the generation of a set of randomly generated solutions (candidate solutions) for a given problem [Abuzairu]. WOA is distinguished from different algorithms by rules that enhance the candidate solutions at every stage of optimization. In reality, WOA imitates the hunting behavior of humpback whales by locating and attacking prey using a technique known as "bubble-net" feeding. LSTM optimization for shorter training durations is a critical problem, particularly when working with large datasets and complex models [Abuzairu].

Data Collection

In this research, Aljarah et al, was used as the primary source of the collected data. (2018) published a journal. Many other journals and websites were also reviewed during data collection. Thus, the study uses six (6) different cancer datasets that are held for bladder, leukemia, lung, ovary, pancreas and prostate. <https://link.springer.com/article/10.1007/s00500-016-2442-1>

Experimental Setup

In all the experiments, we used MATLAB R2021a to implement the proposed WOA trainer and other algorithm. The data is split for 70% training and 30% testing using stratified sampling to preserve the class distribution as best as possible. In addition, all datasets are normalized using min-max normalization to remove the influence of different scaling functions. All experiments are performed in five different runs, where each experiment contains 50, 100, 150, 200, and 250 repetitions. For six (6) different

cancer datasets, the simulation was repeated with 250 iterations in each case.

Parameter Setup

1. Input Size is the argument to the sequence Input Layer function. It is the dimension of the feature, that is, the number of rows in the matrix in each cell.
2. Number of Hidden Units is a parameter of the LSTM Layer function that sets the number of hidden units contained in the LSTM network.
3. Number Class is the argument
4. Of fully Connected Layer, this is the number of labels. For this research, the number of wolves to be identified.

Table 1. Parameter Settings

SN	Parameter	Setting
1	Layer of Sequence Input	Input size
2	The LSTM Layer	2
3	Layer Completely Connected	1
4	Softmax Layer	1
5	Classification Layer	1
6	Max Epochs	7
7	Mini Batch Size	27
8	Gradient Threshold	1
9	Verbose	False
10	Execution Environment	CPU

Evaluation Parameters and Performance Metrics

After implementation, the proposed system will be evaluated based on its performance. Accuracy, convergence speed, and precision score are the performance metrics of this work. These parameters are mathematically computed as follow;

1. Accuracy: this performance metric deals with the correct prediction made by the model and this metric can be expressed as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

2. Precision: precisions provide information about how precise/accurate your model is out of those predicted positives, how many of them actual positives are. Precisions are a good measure to determine when cost of false positives is high. It is mathematically expressed as:

$$Precision = \frac{TP}{TP + FP}$$

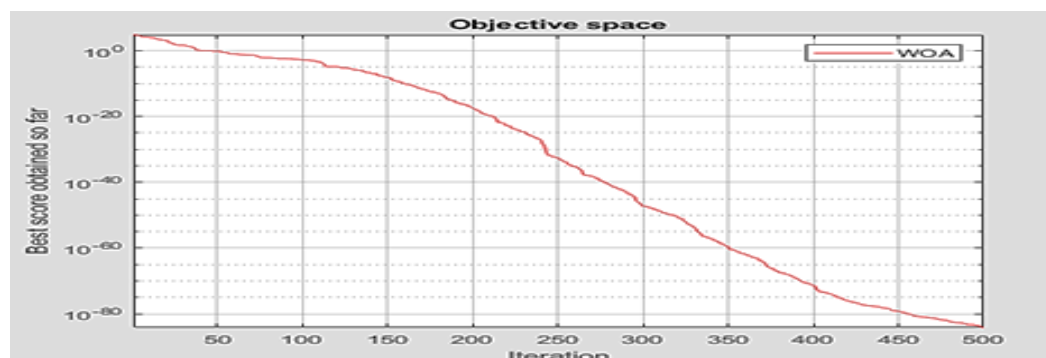


Figure 2: Plots for objective space.

Corresponding author: Abuzairu Ahmad

abuzairuahmad2020@gmail.com

Department of Mathematical Sciences, ATBU, Bauchi.

© 2025. Faculty of Technology Education. ATBU Bauchi. All rights reserved

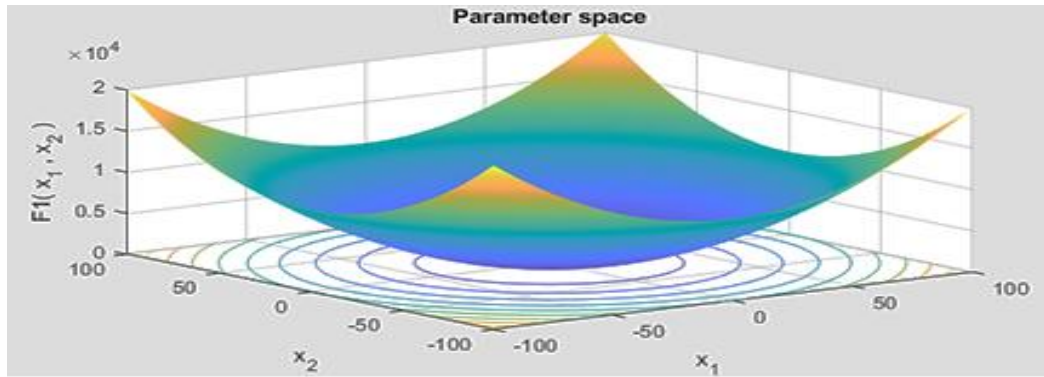


Figure 3: Objective space for whale optimization algorithm.

First, in order to load the specifications of the chosen benchmark function, we chose 30 search agents and 500 iterations as the maximum

number of iterations. As seen in Figure 4, we load the inputs to create the objective space for the whale optimization process.

Table 2: Parameter Settings

SN	Parameter	Setting
1	Layer of Sequence Input	Input size
2	The LSTM Layer	2
3	Layer Completely Connected	1
4	Softmax Layer	1
5	Classification Layer	1
6	Max Epochs	7
7	Mini Batch Size	27
8	Gradient Threshold	1
9	Verbose	False
10	Execution Environment	CPU

RESULTS AND DISCUSSIONS

The proposed WOA trainer is compared to GA in terms of classification accuracy and MSE evaluation metrics. All algorithms were stopped after a predetermined maximum of iterations was reached in order to provide a fair comparison. To provide a comprehensive comparison, the

convergence behavior is also studied in the results. The statistical findings, including mean square error and classification accuracy, as well as the best possible outcome of the proposed WOA, GA are presented in Table 2 for all six (6) different datasets.

Table 3: Results of training and classifications accuracy of the proposed WOA-LSTM against the other compared algorithm used for Bladder, Leukemia, Lung, Ovary, Pancreas and Prostate cancer respectively.

Cancer Datasets	Metric	WOA	GA
Bladder	AVG ACC	98.00	95.00
	AVG MSE	0.02	0.02
	AVG PREC	96.00	95.00
Leukemia	AVG ACC	97.00	94.00
	AVG MSE	0.01	0.01
	AVG PREC	97.00	94.00

Corresponding author: Abuzairu Ahmad

abuzairuhammad2020@gmail.com

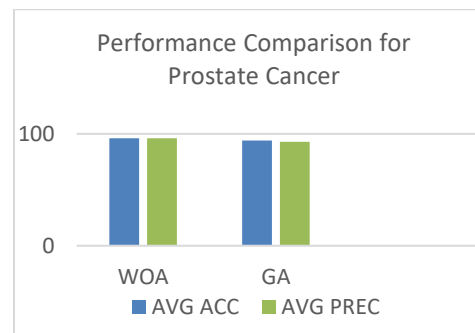
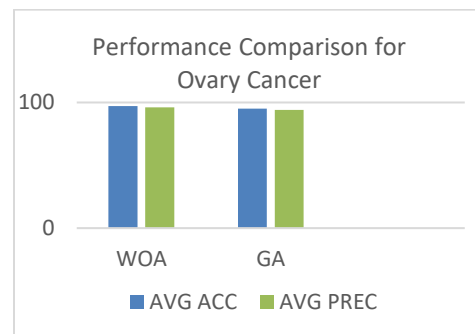
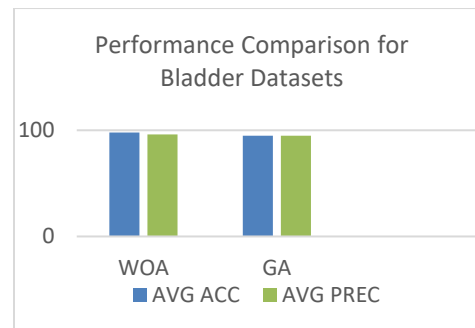
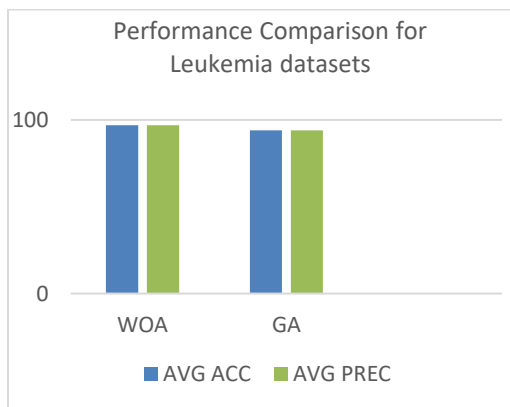
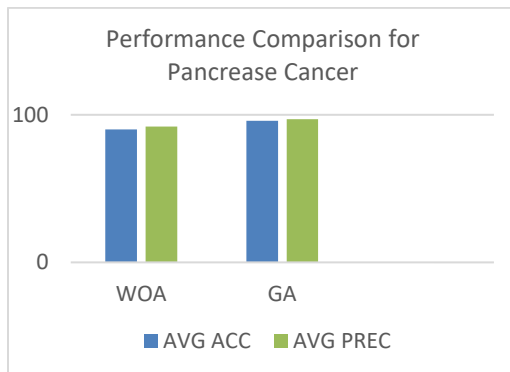
Department of Mathematical Sciences, ATBU, Bauchi.

© 2025. Faculty of Technology Education. ATBU Bauchi. All rights reserved

Cancer Datasets	Metric	WOA	GA
Lung	AVG ACC	87.00	97.00
	AVG MSE	0.01	0.01
	AVG PREC	91.00	99.00
Ovary	AVG ACC	97.00	95.00
	AVG MSE	0.01	0.01
	AVG PREC	96.00	94.00
Pancreas	AVG ACC	90.00	96.00
	AVG MSE	0.01	0.01
	AVG PREC	92.00	97.00
Prostate	AVG ACC	96.00	94.00
	AVG MSE	0.01	0.01
	AVG PREC	96.00	93.00

Performance comparison base on average accuracy and precision

In general, it was noted that the proposed WOA came out on top in four cases (bladder, ovary, leukemia, and prostate cancers), while GA came out on top in two cancer datasets (lung and pancreas cancers).



Corresponding author: Abuzairu Ahmad

abuzairuahmad2020@gmail.com

Department of Mathematical Sciences, ATBU, Bauchi.

© 2025. Faculty of Technology Education. ATBU Bauchi. All rights reserved

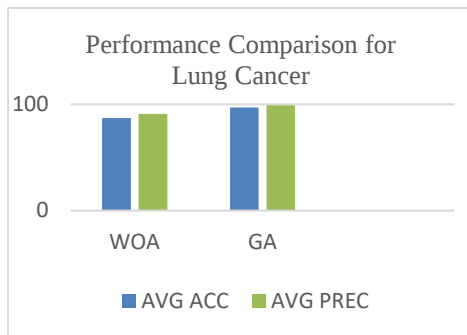


Figure 4: Performance comparison for bladder, leukemia, and pancreas, lung, ovary and prostate cancer

Performance comparison base on convergence speed (MSE)

The proposed WOA's convergence curves, as well as the other algorithm employed in

this research study for assessments, are shown in Figure 7. In general, the suggested WOA on the ovary dataset had the best training MSE of the entire algorithm following the simulation on six (6) distinct datasets, with the lowest MSE of 0.006428. The GA and WOA are calculated using MSE averages for all training samples across five independent runs; optimizers, however, produced the most effective MSE under all other test conditions with improved convergence characteristics. The data indicates that, WOA are the fastest algorithm when compared to GA.

WOA Trainer provides strong evidence that this strategy is capable of reliably preventing premature convergence toward local optima and obtaining the best optimal values for the LSTM's weights and biases. WOA performs very well when compared to the most effective techniques in the majority of scenarios.

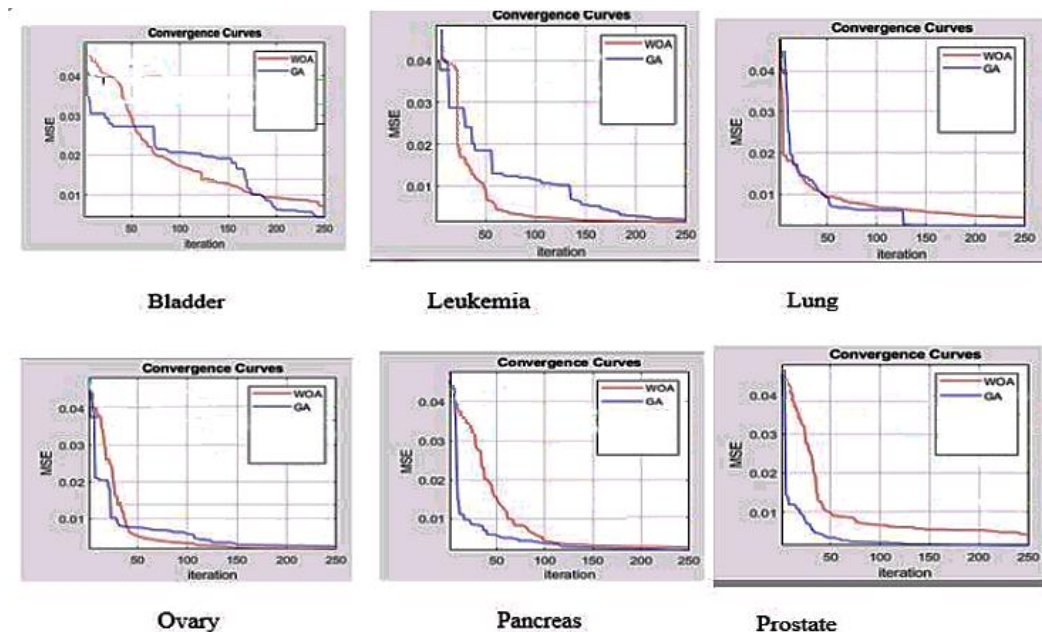


Figure 6: Performance comparison for convergence speed (MSE)

Extensive research has been conducted on LSTM, which has been Classification, pattern recognition, regression, and forecasting problems have all benefited from the use of this technique. The LSTM learning process is widely regarded as one of the most difficult challenges in machine

learning, and it has piqued the interest of several researchers. The primary challenge for researchers in training ANNs (weights and biases) is the algorithm's nonlinearity as well as the unknown best set of primary controlling parameters. Traditional training algorithms have

Corresponding author: Abuzairu Ahmad

abuzairuahmad2020@gmail.com

Department of Mathematical Sciences, ATBU, Bauchi.

© 2025. Faculty of Technology Education. ATBU Bauchi. All rights reserved



two major drawbacks: local optimization stagnation and slow convergence speed. As a result, the stochastic optimization algorithm is a reliable method for mitigating such flaws.

CONCLUSION

In this study, the whale optimization algorithm was proposed to optimize the link weights in a large short-term memory. In addition, the study compares and displays the results of various datasets (bladder, leukemia, lung, ovary, pancreas, and prostate cancer) to determine the best combination of classification accuracy, convergence speed, and precision score. The algorithms described here were written in MATLAB R2021a. In terms of classification accuracy and MSE assessment metrics, the proposed WOA trainer is compared to GA. To ensure a fair comparison, all algorithms were terminated after a set number of iterations. As a final step in providing a comprehensive comparison, the data's convergence behavior is also investigated. The proposed WOA came in first in four of the cases (bladder, ovary, leukemia, and prostate cancers), followed by GA, which came in first in two cancer datasets (lung and pancreas cancers).

In the case of Bladder and Ovary cancer datasets, the proposed model outperforms the other optimizer, with an average MSE of 0.01888 and 0.006428, respectively, whereas the GA was second best, achieving the lowest MSE in Leukemia, Lung, Pancreas, and Prostate Cancers with an average MSE of 0.006984, 0.007946, 0.009627, and 0.008148, respectively. In terms of convergence, the WOA outperformed the other algorithm because of its high exploration and avoidance of local optima. The results also show that increasing local optima avoidance has no effect on WOA convergence. Finally, the proposed trainer has high local optima avoidance, resulting in a short convergence time. Furthermore, RNN can be trained using the proposed trainer for classifying datasets of varying difficulty, as demonstrated in this study.

Another intriguing pattern is the superiority of the Genetic algorithm (GA). This is primarily due to evolutionary algorithms' inherent

higher exploration, which allows them to demonstrate better local optima avoidance. Finally, the WOA-LSTM is more efficient and competitive than WOA-MLP training techniques; it can also train RNNs with a limited or unlimited number of connection weights and biases.

LIMITATIONS

One of the main limitations of this study is that only one variant of RNN could be trained and applied in the study, so for generalization and fair comparison, it is necessary to consider another variant, applying the basic WOA trainer to other versions RNN like Elman Neural Network, Vanilla Recurrent Neural Networks, Gated Recurrent Neural Networks and Non-linear Autoregressive Neural Networks.

RECOMMENDATIONS AND FUTURE WORK

This research is applicable to other deep recurrent neural network variants. As a result, future work will consider training other RNN variants to determine and recommend the best among the variants.

REFERENCES

- Aliyu Aihong A., Imam Ya'u B., Ali U., Ahmad A., Abdulrahman Lawal M. An optimized deep learning method for software defect prediction using Whale Optimization Algorithm. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 2, pp. 222-229. doi: 10.17586/2226-1494-2024-24-2-222-229
- Aljarah, I., Faris, H., & Mirjalili, S. (2018). Optimizing connection weights in neural networks using the whale optimization algorithm. *Soft Computing*, 22(1), 1-15.
- Chui, K. T., Gupta, B. B., & Vasant, P. (2021). A Genetic Algorithm Optimized RNN-LSTM Model for Remaining Useful Life Prediction of Turbofan. *Engine. Electronics*, 10(3), 285.
- Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2020). A survey of deep



- learning and its applications: a new paradigm to machine learning. *Archives of Computational Methods in Engineering*, 27(4), 1071-1092.
- Edoh, T. O. (2020). Reconfiguration Durations Optimization for High-availability Distributed Systems: The case of ICT Rural and Elderly Infrastructures for Development.
- ElSaid, A., Jamiy, F. E., Higgins, J., Wild, B., & Desell, T. (2018, July). Using ant colony optimization to optimize long short-term memory recurrent neural networks. In *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 13-20).
- Gao, Y., Chen, K., Gao, H., Zheng, H., Wang, L., & Xiao, P. (2020). Energy consumption prediction for 3-RRR PPM through combining LSTM neural network with whale optimization algorithm. *Mathematical Problems in Engineering*, 2020.
- Girsang, A. S., Lioexander, F., & Tanjung, D. (2020). Stock price prediction using lstm and search economics optimization. *IAENG International Journal of Computer Science*, 47(4), 758-764. *Journal*, 4(5), 375-383.
- Lei, Y., Li, N., Gontarz, S., Lin, J., Radkowski, S., & Dybala, J. (2018). A model-based method for remaining useful life prediction of machinery. *IEEE Transactions on reliability*, 65(3), 1314-1326.
- Liu, Z. H., Meng, X. D., Wei, H. L., Chen, L., Lu, B. L., Wang, Z. H., & Chen, L. (2021). A regularized LSTM method for predicting remaining useful life of rolling bearings. *International Journal of Automation and Computing*, 18(4), 581-593.
- Ngarambe, J., Irakoze, A., Yun, G. Y., & Kim, G. (2020). Comparative performance of machine learning algorithms in the prediction of indoor daylight illuminances. *Sustainability*, 12(11), 4471.
- Qian, Y., Hu, H., & Tan, T. (2019). Data augmentation using generative adversarial networks for robust speech recognition. *Speech Communication*, 114, 1-9.
- Salman, A. G., Heryadi, Y., Abdurahman, E., & Suparta, W. (2018). Single layer & multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting. *Procedia Computer Science*, 135, 89-98.
- Schmidhuber, J., Gers, F., & Eck, D. (2002). Learning nonregular languages: A comparison of simple recurrent networks and LSTM. *Neural computation*, 14(9), 2039-2041.
- Siarni-Namini, S., & Namin, A. S. (2018). Forecasting economics and financial time series: ARIMA vs. LSTM. *arXiv preprint arXiv:1803.06386*.
- Yadav, A., Jha, C. K., & Sharan, A. (2020). Optimizing LSTM for time series prediction in Indian stock market. *Procedia Computer Science*, 167, 2091-2100.
- Yuliyono, A. D., & Girsang, A. S. (2019). Artificial bee colony-optimized LSTM for bitcoin price prediction. *Advances in Science, Technology and Engineering Systems*
- Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). Dive into deep learning. *arXiv preprint arXiv:2106.11342*.