

Arithmetic Optimization with Deep Learning Enabled Churn Prediction Model for Telecommunication Industries

Vani Haridasan*, Kavitha Muthukumaran and K. Hariharanath

SSN School of Management, Kalavakkam, Chennai, 603110, India

*Corresponding Author: Vani Haridasan. Email: vanih@ssn.edu.in

Received: 30 March 2022; Accepted: 08 June 2022

Abstract: Customer retention is one of the challenging issues in different business sectors, and various firms utilize customer churn prediction (CCP) process to retain existing customers. Because of the direct impact on the company revenues, particularly in the telecommunication sector, firms are needed to design effective CCP models. The recent advances in machine learning (ML) and deep learning (DL) models enable researchers to introduce accurate CCP models in the telecommunication sector. CCP can be considered as a classification problem, which aims to classify the customer into churners and non-churners. With this motivation, this article focuses on designing an arithmetic optimization algorithm (AOA) with stacked bidirectional long short-term memory (SBLSTM) model for CCP. The proposed AOA-SBLSTM model intends to proficiently forecast the occurrence of CC in the telecommunication industry. Initially, the AOA-SBLSTM model performs pre-processing to transform the original data into a useful format. Besides, the SBLSTM model is employed to categorize data into churners and non-churners. To improve the CCP outcomes of the SBLSTM model, an optimal hyperparameter tuning process using AOA is developed. A widespread simulation analysis of the AOA-SBLSTM model is tested using a benchmark dataset with 3333 samples and 21 features. The experimental outcomes reported the promising performance of the AOA-SBLSTM model over the recent approaches.

Keywords: Customer churn prediction; business intelligence; telecommunication industry; decision making; deep learning

1 Introduction

The globalization and developments of telecommunication industries significantly increase the number of workers in the marketplace, intensifying competition [1]. In this competitive period, it is compulsory to maximize the revenues occasionally, for which several approaches have been offered, i.e., obtaining new clients, retaining the prevailing consumers, and raising the retaining duration of prevailing clients [2]. Amongst every strategy, the maintenance of present clients is cheaper when compared to others. Given the implementation of the 3rd policy, businesses need to decrease the significant customer churn (CC), that is, consumer shifts from one service provider to another service provider. The foremost cause for



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.

churn is the disappointment in customer services and support structures. The only way to solve this problem is to predict the clients who are at the risk of churning [3,4].

Nowadays, the challenging environment and cost-cutting stress initiated the organizations to increase the customer relationship management (CRM) module. The subsequent behavior of the clients has to be varied from unfamiliar to familiar ones. However, the estimation of the user's conclusion is very complicated, and earlier findings help prevent the churn prediction [5]. In CRM, user maintenance is a key task that openly affects the company's revenue and production units. For example, a partial raise in the retaining rate has led to the maximization of profits to a greater level. The modification costs relate to users switching nature which is earned by customers, companies, or both. Customers' churn is the primary concern in service-oriented sectors with excellent competitive services [6]. In addition, forecasting the consumers who are probably to terminate the firm will denote a potentially huge additional income source if it is performed at the initial stage. Several researchers assured that machine learning (ML) knowledge effectively forecasts this condition. This technique is practiced by learning from existing data [7].

The existence of digital systems arrangement, information technology, ML, statistical, and existing methods provide improved chances for telecommunication companies to study the user necessities that are necessary to show proficient CRM [8]. The financial state is affected by the users in the telecommunication industries, so even a minor alteration in financial objective has to be keenly watched and scrutinized. Here, CC is caused by the changes made in user behavior. However, identifying client opinions is impossible even by implementing tactics and systematic concepts [9]. It is mandatory to execute a novel study to learn consumer characteristics to identify the new aspects of CC. Employing CRM becomes an advantage in efficiently reducing the churn rate. ML has verified itself as a highly effective approach for predicting information based on earlier captured data [10] to overcome this issue.

The researcher in [10] executes the customer churn prediction (CCP) based on the assumption that the influencer has passionate support from the follower. The information gathered by the influencer marketing agency in Korea from August 2018 to October 2020 comprises the purchase details: purchase item, payment amount, and customer information. To forecast the churning customer, we employ the Decision Tree (DT) approach. Zhao et al. [11] analyzed the causes and trends of CC via a data mining algorithm. They provided the answer to this question as to how enterprises win back churned customers, how the CC takes place, and the influencing factors of CC. The experimental result serves the practice of CRM better in the telecom industry and references the telecom industry to recognize higher-risk churned customers.

In [12], a class-dependent cost-sensitive boosting approach, AdaBoostWithCost, is presented to reduce the churn cost. It illustrates the empirical evaluation of the proposed method that reliably outperforms the discrete AdaBoost approach concerning telecom churn prediction. Zhang et al. [13] attempted to develop a churn prediction method for predicting telecom client churn via customer segmentation. Information is gathered from three Chinese telecom corporations, and Fisher discriminatory equation and logistic regression analysis have been utilized for building a telecom CCP method [14]. Based on the outcomes, it could be concluded that the telecom CC method built by regression analysis had a high prediction performance.

This article focuses on designing an arithmetic optimization algorithm (AOA) with stacked bidirectional long short-term memory (SBLSTM) model for CCP. The proposed AOA-SBLSTM model intends to proficiently forecast the occurrence of CC in the telecommunication industry. Initially, the AOA-SBLSTM model performs pre-processing to transform the original data into a useful format. Besides, the SBLSTM model is employed to categorize data into churners and non-churners. To improve the CCP outcomes of the SBLSTM model, an optimal hyperparameter tuning process using AOA is developed. A

widespread simulation analysis of the AOA-SBLSTM model is tested using a benchmark dataset with 3333 samples and 21 features.

2 The Proposed CCP Model

In this study, a novel AOA-BLSTM model has been designed to proficiently forecast the occurrence of CC in the telecommunication industry. Initially, the AOA-SBLSTM model performs pre-processing to transform the original data into a useful format. Besides, the SBLSTM model is employed to categorize data into churners and non-churners. To improve the CCP outcomes of the SBLSTM model, an optimal hyperparameter tuning process using AOA is developed. Fig. 1 demonstrates the overall process of the AOA-BLSTM technique.

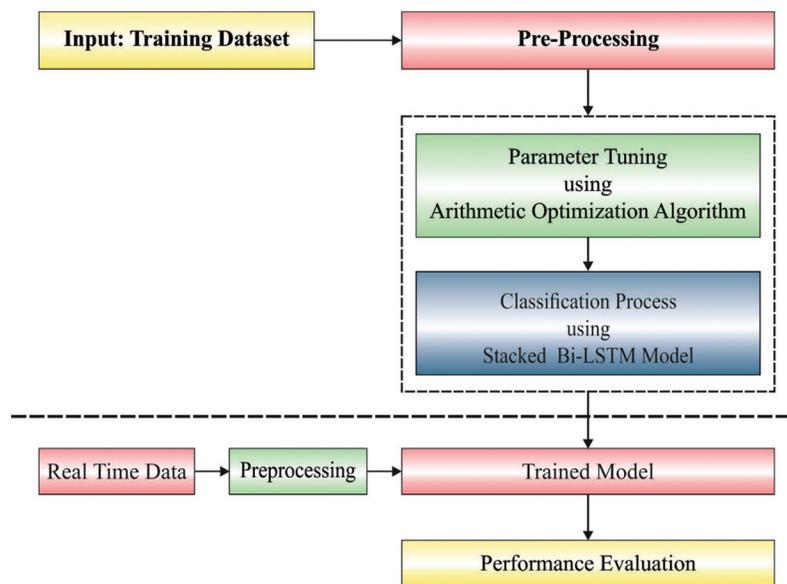


Figure 1: The overall process of the AOA-SBLSTM technique

2.1 Data Pre-processing

Firstly, the AOA-SBLSTM model follows z-score normalization-based pre-processing to transform the original data into a useful format. The Z-score is evaluated by the arithmetic mean, and the standard deviation of the input data is commonly employed as the score normalization method. Predictably, this normalization method performs well in the case of previous knowledge about the score variation and the average score of the matcher. The normalized score is represented as follows

$$s'_k = \frac{s_k - \mu}{\sigma} \quad (1)$$

whereas σ denotes the standard deviation and μ indicates the arithmetic mean of the data. In our work, the normalization of the smoothed data is performed by Z-score normalization.

2.2 SBLSTM Based Prediction

At this stage, the SBLSTM model is employed to categorize data into churners and non-churners. The normal structure of LSTM efficiently resolves the gradient vanishing problems and transports valuable data within the LSTM network [15]. The Recurrent Neural Network (RNN) failed to capture long-term

dependency amongst feature vectors. The LSTM cell comprises of input gate (i_t), forget gate (f_t), and output gate (o_t). That gate controls a memory cell activation vector.

- Forget gate: it can be utilized for determining the number of data from a preceding layer c_{t-1} that it can be utilized for determining the number of forgotten or retained according to the existing input x_t and hidden layer h_{t-1} . The forget gate can be formulated below.

$$f_t = \text{Sigmoid}(w_{xf}x_t + w_{hf}h_{t-1} + b_f) \quad (2)$$

whereas b_f , w_{xf} , and w_{hf} denotes the bias vector; the weighted matrix is among x_t and f_t , and the weighted matrix is among h_{t-1} and f_t , correspondingly

- Input gate: it can be utilized for determining the number of the network's input x_t that need to be retained in the existing cell state c_t .

$$i_t = \text{Sigmoid}(w_{xi}x_t + w_{hi}h_{t-1} + b_i) \quad (3)$$

In which b_i , w_{xi} , and w_{hi} denotes bias vector, the weighted matrix is among x_t and i_t , and the weighted matrix is among h_{t-1} and i_t , correspondingly.

- Output gate: it can be utilized for determining the number of data transported to the LSTM in the cell state c_t via the present output h_t . LSTM gate is a completely connected network where input represents a vector and output denotes a real number. The output gate can be formulated below.

$$o_t = \text{Sigmoid}(w_{xo}x_t + w_{ho}h_{t-1} + b_o) \quad (4)$$

whereas b_o , w_{xo} , and w_{ho} denotes the bias vectors, the weight matrix is among x_t and o_t , also the weight matrix is among h_{t-1} and o_t , correspondingly

The last output of the LSTM cell is the cell output state c_t and layer output h_t , that is formulated below, correspondingly

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

The intermediate cell input is denoted as \tilde{c}_t that is formulated by:

$$\tilde{c}_t = \tanh(w_{xc}x_t + w_{hc}h_{t-1} + b_c) \quad (7)$$

Here b_c , w_{xc} , and w_{hc} indicates the bias vector, the weight matrix is among x_t and \tilde{c}_t , and the weight matrix is among h_{t-1} and \tilde{c}_t correspondingly. Fig. 2 illustrates the infrastructure of Stacked Bidirectional Long Short-Term Memory (BiLSTM) [15].

The LSTM is employed by fruit fly optimization for NN3 information for time series analysis. Alternatively, LSTM was utilized [16] in a fully varied field to forecast the lifetime of equipment in the mechanical broadcast scheme. A BiLSTM encompasses two distinct LSTMs that combine data from two directions. Then, the data attained as word annotation from the customer data is combined. The forward LSTM process uses the input from left to right and estimates the hidden layer \vec{h}_t according to x_t and \vec{h}_{t-1} , whereas the backward LSTM process uses the input from right to left and estimates the hidden layer \overleftarrow{h}_t according to x_t and \overleftarrow{h}_{t-1} . The forward and backward variables in the BiLSTM are autonomous of each other's, even though word embedding is distributed. At last, the hidden layer of BiLSTM is determined that concatenate the vector of the forward and backward directions during time step t .

$$h_t = \left[\vec{h}_t, \overleftarrow{h}_t \right] \quad (8)$$

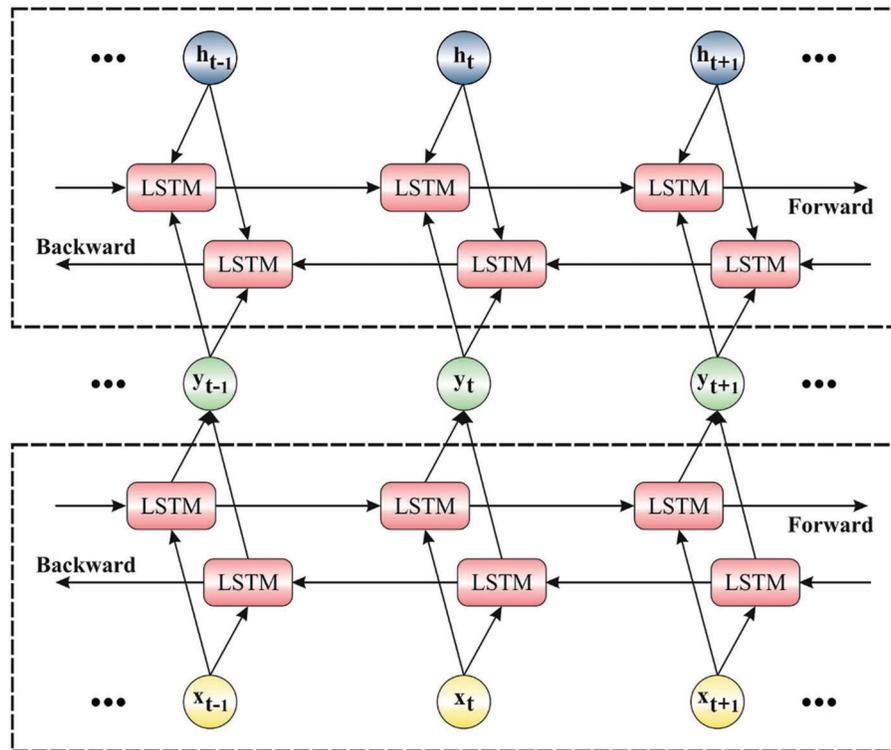


Figure 2: Structure of stacked BiLSTM

Further, the deep hierarchical method improves the accuracy of the classification. But the stacked BiLSTM is very effective in the shallow learning method. Hence the presented method determines a stacking BiLSTM to employ the local context and latent symmetry complex data. The output of the low layer becomes the input of the upper layer in a Stacked BiLSTM.

2.3 AOA Based Hyperparameter Tuning

At the final stage, an optimal hyperparameter tuning process using AOA is developed to improve the performance of the SBLSTM model. AOA is a novel meta-heuristic model presented by Abualigah and others in 2021 [16]. The main stimulus of this technique is for combining the four standard arithmetic operators from mathematics, namely division (D), multiplication (M), addition (A), and subtraction (S). Related to the sine-cosine algorithm (SCA), AOA is a straightforward infrastructure and has minimal computation complexity. Assuming M and D operators are producing huge stages from the iterations, M and D are, therefore, mostly conducted from the exploration stage. The formulation is as follows:

$$Xi(t + 1) = \begin{cases} Xb(t)/(MOP + eps) \cdot ((UB - LB)\mu + LB) , rand < 0.5 \\ Xb(t) \cdot MOP \cdot ((UB - LB)\mu + LB) , rand \geq 0.5 \end{cases} \quad (9)$$

whereas eps refers to the very easy positive number, and μ signifies the constant co-efficient (0.499), which is carefully planned for this approach.

In MOP, it is non-linearly reduced from one to zero under the iterations, and the formulation is as follows:

$$MOP = 1 - \left(\frac{t}{T}\right)^{1/\alpha} \quad (10)$$

In which α refers to the constant value fixed to 5 based on the AOA.

In Eq. (9), it could be noticeable that both M and D operators make very stochastic places to search agents on the fundamental of optimum place. Conversely, S and A operators were executed to emphasize the local exploitation, creating lesser stages under the searching space. The mathematical formula is determined as:

$$X_i(t+1) = \begin{cases} X_b(t) - MOP \cdot ((UB - LB)\mu + LB), & rand < 0.5 \\ X_b(t) + MOP \cdot ((UB - LB)\mu + LB), & rand \geq 0.5 \end{cases} \quad (11)$$

. There is no doubt that the balance between exploration and exploitation is significant to an optimized technique. In AOA, the parameter MOA was employed for switching the exploration as well as exploitation on the course of iterations that is written as:

$$MOA(t) = Min + t \left(\frac{Max - Min}{T} \right) \quad (12)$$

whereas Min and Max are constant values.

Based on Eq. (12), MOA improves in Min to Max. Therefore, under the primary stage, the searching agent is a further chance for performing exploration from the searching space, but under the later phase, the searching agent is highly possible for conduct searching neighboring the optimum place. The pseudo-code of AOA is demonstrated in Algorithm 1.

The AOA approach derives a fitness function to obtain better classification accuracy. It defines a positive integer to characterize the effective accuracy of the candidate solutions. In our work, the minimization of the classification error rate is regarded as the fitness function. The optimum solution has a small error rate, and the worse solution accomplishes an increased error rate.

$$fitness(x_i) = ClassifierErrorRate(x_i) = \frac{number\ of\ misclassified\ samples}{Total\ number\ of\ samples} * 100 \quad (13)$$

Algorithm 1: Pseudo-code of AOA

Initializing the parameters popsize (N) and maximal iterations (T)

Initializing the places of every searching agent $X_i(i = 1, 2, \dots, N)$

Set the parameters α , μ , Min, and Max

While ($t \leq T$)

Compute the fitness of every search agent Upgrade optimum Fitness, X_b

Compute the MOP by Eq. (10)

Compute the MOA by Eq. (12)

For all search agents

If $rand > MOA$

Upgrade place by Eq. (9)

Else

Upgrade place by Eq. (11)

End if

End for

$t = t + 1$

End While

Return optimum Fitness, X_b

3 Results and Discussion

In this section, the experimental validation of the AOA-SBLSTM model is tested using a benchmark CCP dataset (available at <http://www.sgi.com/tech/mlc/db/>), comprising 3333 instances with 21 features. A set of 483 samples falls into the churner (CR) class, and 2859 samples come under the non-churner (NCR) class. The proposed model is simulated using Python 3.6.5 tool.

Fig. 3 inspects the performance of the AOA-SBLSTM model under distinct epoch count. The results demonstrated that the AOA-SBLSTM model has effectively classified CR and NCR classes. For instance, with 200 epochs, the AOA-SBLSTM model has identified 342 and 2844 samples under CR and NCR classes, respectively. In addition, with 400 epochs, the AOA-SBLSTM methodology has identified 208 and 2845 samples under CR and NCR classes correspondingly. Moreover, with 800 epochs, the AOA-SBLSTM technique has correspondingly identified 346 and 2843 samples under CR and NCR classes. Furthermore, with 1000 epochs, the AOA-SBLSTM system has correspondingly identified 310 and 2844 samples under CR and NCR classes. With 1200 epochs, the AOA-SBLSTM technique has correspondingly identified 371 and 2843 samples under CR and NCR classes.

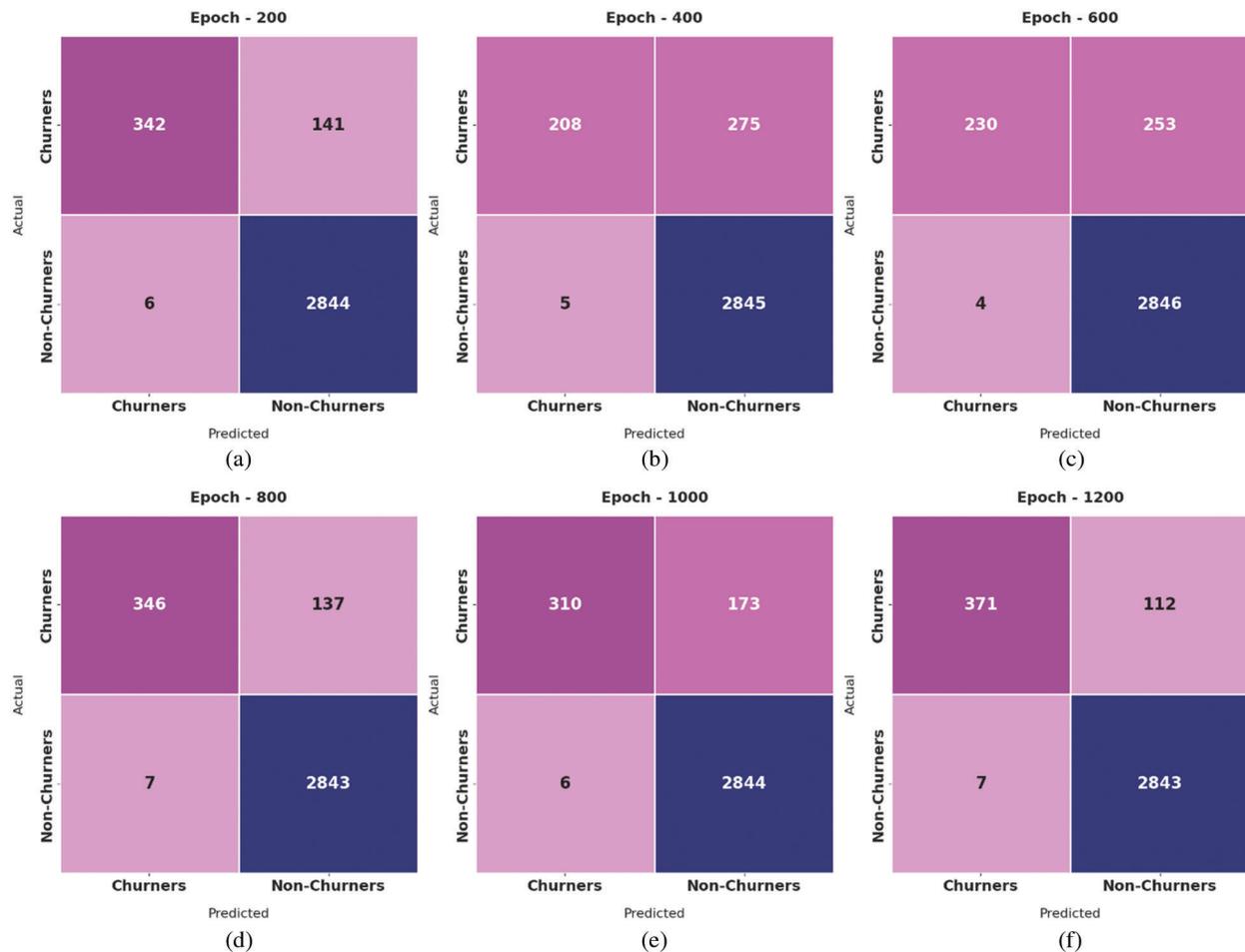


Figure 3: Confusion matrix of AOA-SBLSTM technique under distinct epoch count

Tab. 1 and Fig. 4 highlights the overall CCP outcomes of the AOA-SBLSTM model under varying epoch count. The experimental values denoted the improved performance of the AOA-SBLSTM model

under all epochs. For instance, with 100 epochs, the AOA-SBLSTM model has obtained an average $accu_y$, $prec_n$, $reca_l$, $spec_y$, F_{score} , and Mathew Correlation Coefficient (MCC) of 95.59%, 96.78%, 85.30%, 85.30%, 89.90%, and 81.27% respectively. Moreover, with 400 epochs, the AOA-SBLSTM system has reached an average $accu_y$, $prec_n$, $reca_l$, $spec_y$, F_{score} , and MCC of 91.60%, 94.42%, 71.44%, 71.44%, 77.54%, and 61.73% correspondingly. Eventually, with 800 epochs, the AOA-SBLSTM technique has achieved an average $accu_y$, $prec_n$, $reca_l$, $spec_y$, F_{score} , and MCC of 95.68%, 96.71%, 85.69%, 85.69%, 90.15%, and 81.67% respectively. Simultaneously, with 1000 epochs, the AOA-SBLSTM methodology has obtained an average $accu_y$, $prec_n$, $reca_l$, $spec_y$, F_{score} , and MCC of 94.63%, 96.18%, 81.99%, 81.99%, 87.27%, and 76.87% correspondingly. Concurrently, with 1200 epochs, the AOA-SBLSTM technique has reached an average $accu_y$, $prec_n$, $reca_l$, $spec_y$, F_{score} , and MCC of 96.43%, 97.18%, 88.28%, 88.28%, 92.06%, and 85% correspondingly.

Table 1: Result analysis of AOA-SBLSTM technique with distinct measures

Class labels	Accuracy	Precision	Recall	Specificity	F-score	MCC
Epcoh-200						
Churners	95.59	98.28	70.81	99.79	82.31	81.27
Non-churners	95.59	95.28	99.79	70.81	97.48	81.27
Average	95.59	96.78	85.30	85.30	89.90	81.27
Epcoh-400						
Churners	91.60	97.65	43.06	99.82	59.77	61.73
Non-Churners	91.60	91.19	99.82	43.06	95.31	61.73
Average	91.60	94.42	71.44	71.44	77.54	61.73
Epcoh-600						
Churners	92.29	98.29	47.62	99.86	64.16	65.41
Non-churners	92.29	91.84	99.86	47.62	95.68	65.41
Average	92.29	95.06	73.74	73.74	79.92	65.41
Epcoh-800						
Churners	95.68	98.02	71.64	99.75	82.78	81.67
Non-churners	95.68	95.4	99.75	71.64	97.53	81.67
Average	95.68	96.71	85.69	85.69	90.15	81.67
Epcoh-1000						
Churners	94.63	98.10	64.18	99.79	77.60	76.87
Non-churners	94.63	94.27	99.79	64.18	96.95	76.87
Average	94.63	96.18	81.99	81.99	87.27	76.87
Epcoh-1200						
Churners	96.43	98.15	76.81	99.75	86.18	85.00
Non-churners	96.43	96.21	99.75	76.81	97.95	85.00
Average	96.43	97.18	88.28	88.28	92.06	85.00

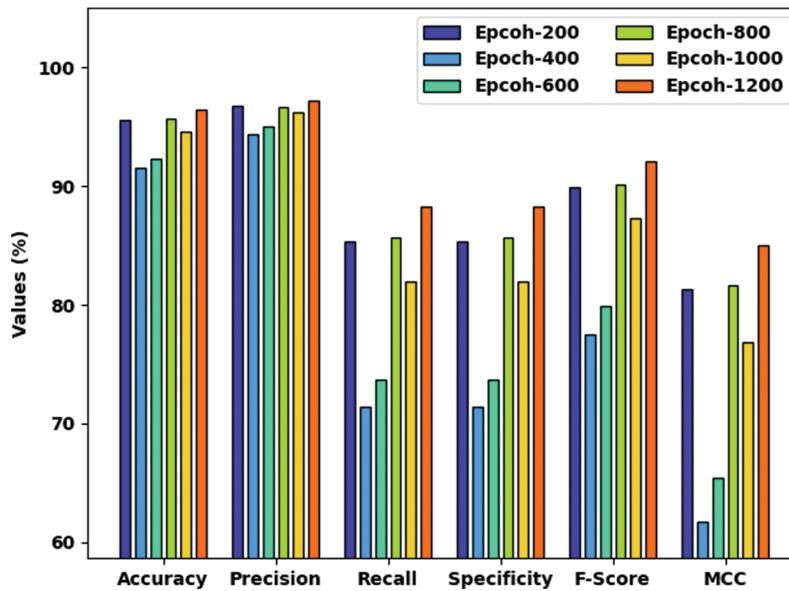


Figure 4: Result analysis of AOA-SBLSTM technique with distinct measures

The training accuracy (TA) and validation accuracy (VA) attained by the AOA-SBLSTM approach on the test dataset is demonstrated in Fig. 5. The experimental outcomes implied that the ICSOA-DLPEC model has gained maximal values of TA and VA. In specific, the VA seemed to be higher than TA.

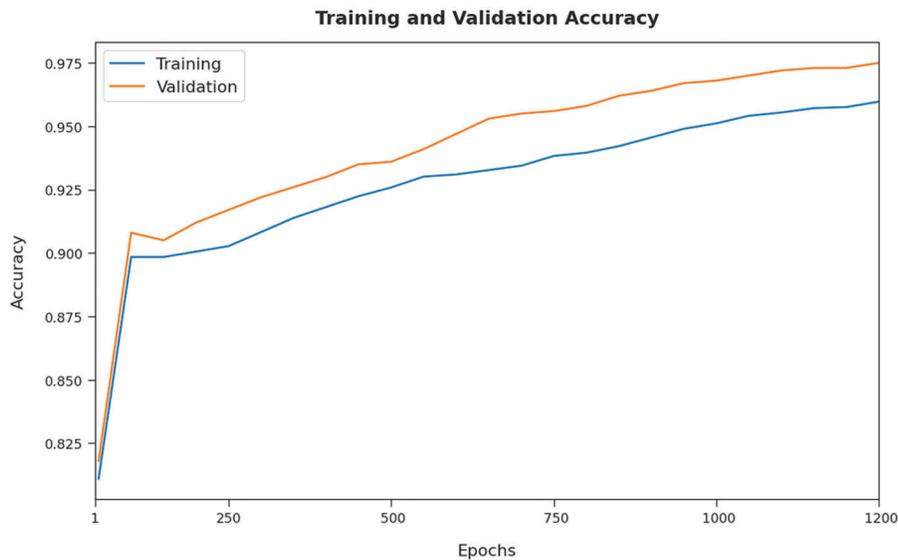


Figure 5: TA and VA graph analysis of AOA-SBLSTM technique

The training loss (TL) and validation loss (VL) attained by the AOA-SBLSTM technique on the test dataset are established in Fig. 6. The experimental outcomes inferred that the ICSOA-DLPEC system had accomplished the least values of TL and VL. In specific, the VL seemed to be lower than TL.

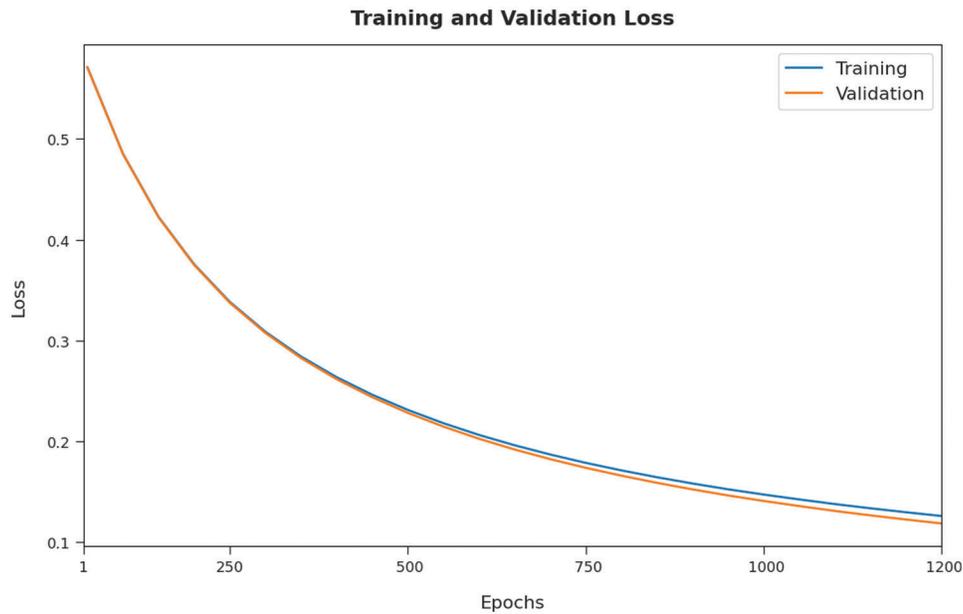


Figure 6: TL and VL graph analysis of AOA-SBLSTM technique

A detailed precision-recall examination of the AOA-SBLSTM model on the test dataset is demonstrated in Fig. 7. By observing Fig. 7, it can be obvious that the AOA-SBLSTM model has accomplished maximal precision-recall performance under the test dataset.

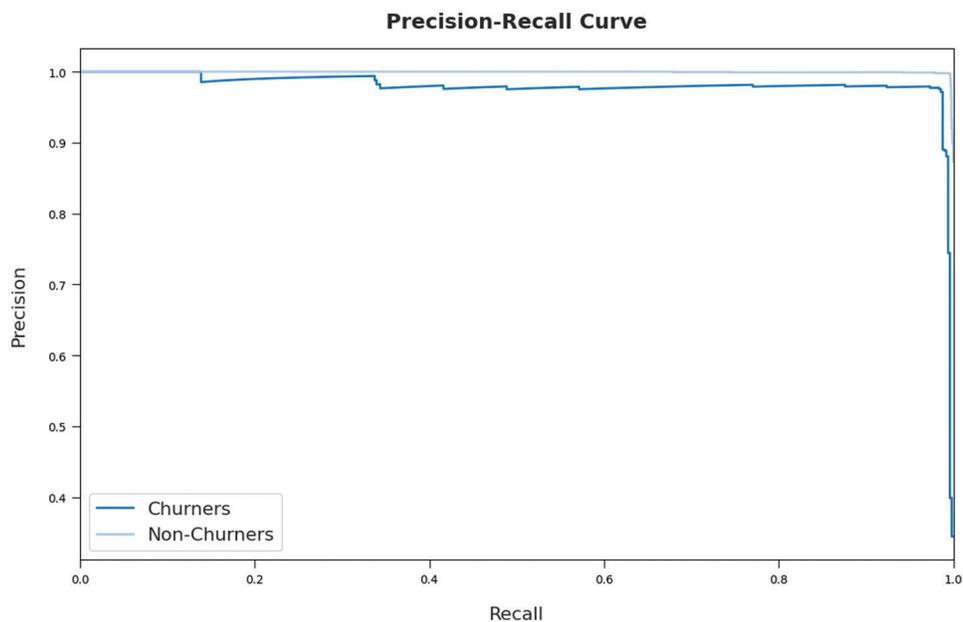


Figure 7: Precision-recall curve analysis of AOA-SBLSTM technique

Fig. 8 depicts the Receiver operating characteristic (ROC) examination of the AOA-SBLSTM model on the test dataset. Fig. 8 exposed that the AOA-SBLSTM model has resulted in maximal ROC values on the categorization of churners and non-churners images.

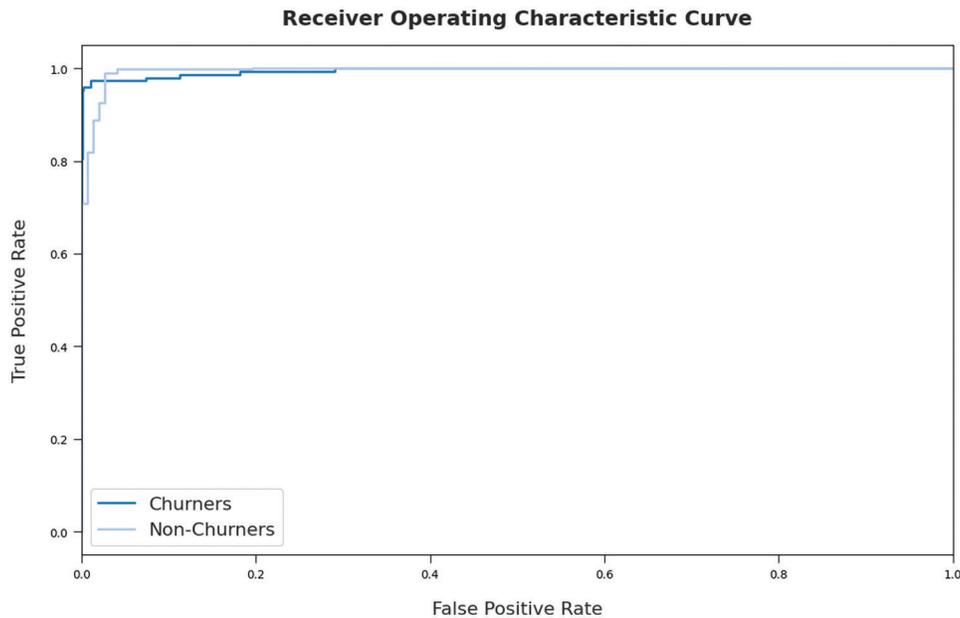


Figure 8: ROC curve analysis of AOA-SBLSTM technique

In order to ensure the enhancements of the AOA-SBLSTM model, a comparative study with existing approaches [16,17] is provided in Tab. 2. Fig. 9 reports a comparative $accu_y$ and $prec_n$ examination of the AOA-SBLSTM model with recent models. The results exhibited that the extreme gradient boosting (XGBoost) model has showcased lower values of $accu_y$ and $prec_n$. At the same time, the Adaboost and Random Forest models have obtained slightly enhanced values of $accu_y$ and $prec_n$. In line with this, the PCPM model has accomplished reasonable values of $accu_y$ and $prec_n$. Though the optimal weighted extreme learning machine (OWELM) and weighted extreme learning machine (WELM) models have resulted in considerable values of $accu_y$ and $prec_n$, the AOA-SBLSTM model has gained maximum performance with $accu_y$ and $prec_n$ of 96.43% and 97.18% respectively.

Table 2: Comparative analysis of AOA-SBLSTM technique with recent algorithms

Methods	Accuracy	Precision	Recall	F-score
Adaboost algorithm	81.71	81.21	80.14	80.28
Random forest	81.21	81.28	80.19	80.28
XGBoost algorithm	80.80	80.70	80.30	78.70
OWELM algorithm	89.24	91.38	87.54	87.12
WELM algorithm	89.89	89.18	85.30	90.15
PCPM algorithm	86.59	89.46	87.37	87.19
AOA-SBLSTM	96.43	97.18	88.28	92.06

Fig. 10 demonstrates a comparative $reca_l$ and F_{score} analysis of the AOA-SBLSTM algorithm with existing methods. The results outperformed that the XGBoost technique has showcased lesser values of $reca_l$ and F_{score} . Besides, the Adaboost and Random Forest approaches have reached somewhat enhanced values of $reca_l$ and F_{score} . Also, the PCPM model has accomplished reasonable values of $reca_l$ and F_{score} .

But, the OWELM and WELM models have resulted in considerable values of $reca_l$ and F_{score} , the AOA-SBLSTM technique has gained maximum performance with $reca_l$ and F_{score} of 88.28% and 92.06% correspondingly.

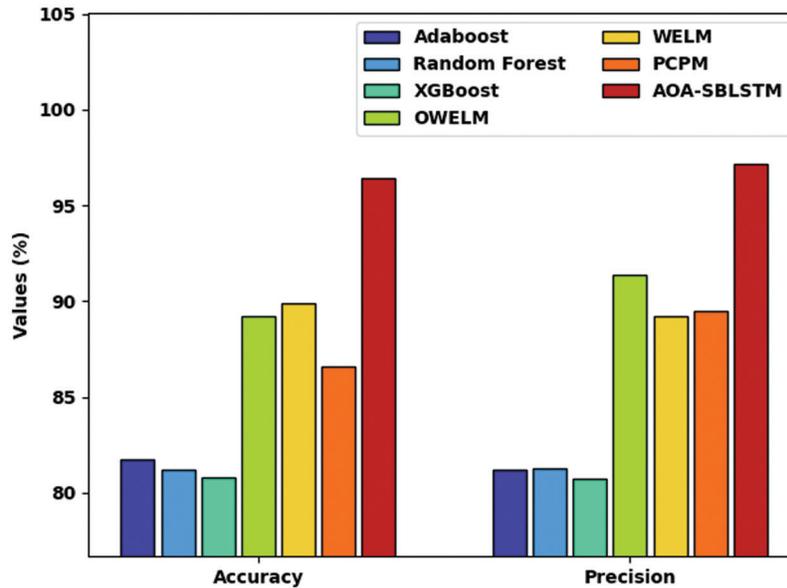


Figure 9: Acc_y and $Prec_n$ analysis of the AOA-SBLSTM technique with existing methodologies

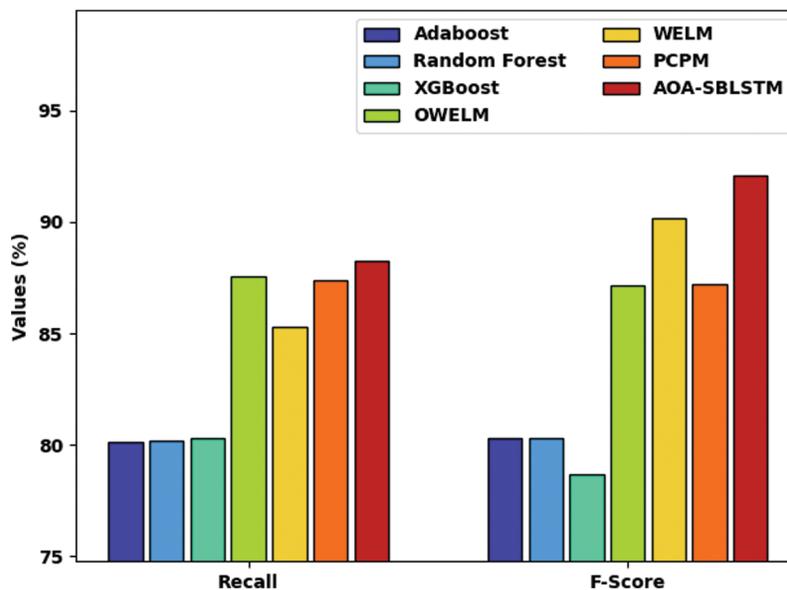


Figure 10: $Recal_l$ and F_{score} analysis of the AOA-SBLSTM technique with existing methodologies

After observing [Tab. 2](#), [Figs. 9](#) and [10](#), it is apparent that the AOA-SBLSTM model has shown an effectual outcome on the CCP process.

4 Conclusion

In this study, a novel AOA-BLSTM model has been designed to proficiently forecast the occurrence of CC in the telecommunication industry. Initially, the AOA-SBLSTM model performs pre-processing to transform the original data into a useful format. Besides, the SBLSTM model is employed to categorize data into churners and non-churners. To improve the CCP outcomes of the SBLSTM model, an optimal hyperparameter tuning process using AOA is developed. A widespread simulation analysis of the AOA-SBLSTM model is tested using a benchmark dataset with 3333 samples and 21 features. The experimental outcomes reported the promising performance of the AOA-SBLSTM model over the recent approaches. Thus, the presented AOA-SBLSTM model can be exploited as a real-time CCP tool. In the future, feature selection methodologies can be incorporated to reduce complexity and increase prediction efficiency.

Funding Statement: The authors received no specific funding for this study.

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

References

- [1] A. K. Ahmad, A. Jafar and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," *Journal of Big Data*, vol. 6, no. 1, pp. 1–24, 2019.
- [2] A. De Caigny, K. Coussement, K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *European Journal of Operational Research*, vol. 269, no. 2, pp. 760–772, 2018.
- [3] A. Amin, F. Al-Obeidat, B. Shah, A. Adnan, J. Loo *et al.*, "Customer churn prediction in telecommunication industry using data certainty," *Journal of Business Research*, vol. 94, no. 8, pp. 290–301, 2019.
- [4] J. Uthayakumar, N. Metawa, K. Shankar and S. K. Lakshmanaprabu, "Financial crisis prediction model using ant colony optimization," *International Journal of Information Management*, vol. 50, no. 5, pp. 538–556, 2020.
- [5] F. Shirazi and M. Mohammadi, "A big data analytics model for customer churn prediction in the retiree segment," *International Journal of Information Management*, vol. 48, no. 6, pp. 238–253, 2019.
- [6] A. De Caigny, K. Coussement, K. W. De Bock and S. Lessmann, "Incorporating textual information in customer churn prediction models based on a convolutional neural network," *International Journal of Forecasting*, vol. 36, no. 4, pp. 1563–1578, 2020.
- [7] P. Lalwani, M. K. Mishra, J. S. Chadha and P. Sethi, "Customer churn prediction system: A machine learning approach," *Computing*, vol. 104, pp. 1–24, 2021.
- [8] E. Stripling, S. vanden Broucke, K. Antonio, B. Baesens and M. Snoeck, "Profit maximizing logistic model for customer churn prediction using genetic algorithms," *Swarm and Evolutionary Computation*, vol. 40, no. 3, pp. 116–130, 2018.
- [9] S. De, P. Prabu and J. Paulose, "Effective ML techniques to predict customer churn," in *2021 Third Int. Conf. on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, IEEE, pp. 895–902, 2021.
- [10] S. Kim and H. Lee, "Customer churn prediction in influencer commerce: An application of decision trees," *Procedia Computer Science*, vol. 199, no. 2, pp. 1332–1339, 2022.
- [11] M. Zhao, Q. Zeng, M. Chang, Q. Tong and J. Su, "A prediction model of customer churn considering customer value: An empirical research of telecom industry in China," *Discrete Dynamics in Nature and Society*, vol. 2021, no. 5, pp. 1–12, 2021.
- [12] H. K. Thakkar, A. Desai, S. Ghosh, P. Singh and G. Sharma, "Clairvoyant: AdaBoost with cost-enabled cost-sensitive classifier for customer churn prediction," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, pp. 1–11, 2022.
- [13] T. Zhang, S. Moro and R. F. Ramos, "A data-driven approach to improve customer churn prediction based on telecom customer segmentation," *Future Internet*, vol. 14, no. 3, pp. 94, 2022.

- [14] A. K. Nandanwar and J. Choudhary, "Semantic features with contextual knowledge-based web page categorization using the GloVe model and stacked BiLSTM," *Symmetry*, vol. 13, no. 10, pp. 1772, 2021.
- [15] I. Banerjee, Y. Ling, M. C. Chen, S. A. Hasan, C. P. Langlotz *et al.*, "Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification," *Artificial Intelligence in Medicine*, vol. 97, no. 2, pp. 79–88, 2019.
- [16] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz and A. H. Gandomi, "The arithmetic optimization algorithm," *Computer Methods in Applied Mechanics and Engineering*, vol. 376, no. 2, pp. 113609, 2021.
- [17] I. V. Pustokhina, D. A. Pustokhin, P. T. Nguyen, M. Elhoseny and K. Shankar, "Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector," *Complex & Intelligent Systems*, pp. 1–13, 2021.