IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 17 July 2023, accepted 8 August 2023, date of publication 14 August 2023, date of current version 5 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3304669

RESEARCH ARTICLE

Arithmetic Optimization With Ensemble Deep Learning SBLSTM-RNN-IGSA Model for Customer Churn Prediction

NAGARAJU JAJAM[1], NAGENDRA PANINI CHALLA[1], (Member, IEEE), KAMEPALLI S. L. PRASANNA[1], AND C H VENKATA SASI DEEPTHI[2]

[1]School of Computer Science and Engineering (SCOPE), VIT-AP University, Amaravati, Andhra Pradesh 522237, India
[2]Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women (Autonomous), Bhimavaram 534202, India

Corresponding author: Nagendra Panini Challa (nagendra.challa@vitap.ac.in)

ABSTRACT Companies in a wide variety of industries use the customer churn prediction (CCP) process to keep their current clientele happy. Insurance companies need to be able to forecast churn to enhance the potency and functionality of deep learning methods. Deep learning techniques have a significant impact on improving and forecasting customer retention. Numerous studies employ standard machine learning and Deep Learning strategies to enhance customer retention, despite the fact that these strategies have a number of accuracy issues. In light of this need, this piece is dedicated to the development of a stacked bidirectional long short-term memory (SBLSTM) and RNN model for Arithmetic Optimisation Algorithm (AOA) in CCP. The proposed AOA-SBLSTM-RNN model intends to proficiently forecast the occurrence of Customer Churn in the Insurance industry. Initially, the AOA model performs pre-processing to transform the original data into a useful format. In addition, the SBLSTM-RNN model is used to distinguish between churning and non-churning customers. To improve the CCP outcomes of the SBLSTM-RNN model, an optimal Hyperparameters tuning process using Improved Gravitational Search Optimization Algorithm (IGSA) is used in this study. In this work, Three Health Insurance datasets were used to evaluate performance, and four sets of experiments were conducted. The Measures of true churn, false churn, specificity, precision, and accuracy are employed to assess the efficacy of the proposed approach. Experimental result shows that the Ensemble Deep Learning model AOA-SBLSTM-RNN with IGSA produces accuracy value of 97.89 and 97.67 on dataset 2 and dataset 1. which is better and had higher predictability levels in compared with all other models.

INDEX TERMS Customer churn, insurance industry, ensemble deep learning, arithmetic optimization algorithm, feature extraction, stacked bidirectional long short-term memory, RNN, improved gravitational search optimization algorithm.

I. INTRODUCTION

The Insurance sector is one of the biggest and most visible users of data forecasting technologies. This theory takes into account insurers' heavy reliance on data and their ever-expanding customer base. This is a daunting concept for both academics and businesses. Clients can now switch insurance providers, indicating that the insurance company's churn probability prediction isn't reaching their needs. Thus, we discuss churn forecasting with ML and DL methods. The financial industry is large and intricate, so it adapts quickly to changes in the global economy. Researchers with an interest in finding practical solutions to problems will find the development of a reliable and honest customer relationship management (CRM) system to be an intriguing topic. [1], [2], [3].

Why Churn prediction: Summarise the significance of actions involving churn reduction and client retention in five main points:

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

Retention of clients minimizes the need for firms to advertise for new clients, allowing them to focus on developing existing connections.

- ✓ Older clients, who are interested in the firms, and desire to buy even more, if fulfilled, can refer others.
- ✓ Because of the enhanced information gathered during their useful life span, engaging and retaining long-term customers is less expensive.
- ✓ Customers who stay with you for a long time are usually more beneficial.
- ✓ Customer loss is an additional expense to firms. it minimizes profits of the gaining of new consumers to offset losses [4].

Many studies are attempting to construct a model that can accurately forecast customer churn using deep learning algorithms. However, present deep learning models still have room for improvement in terms of accuracy. This is where our work in constructing a deep learning model capable of forecasting churn rate with good accuracy when compared to conventional models and dealing with large volumes of data comes into play [5]. Professionals in the field of marketing estimate that annually, companies lose 26% of their customers due to inefficiency and the inability to search for a specific item. The rate at which customers are ready to cancel their contracts and switch to a competitor is a major issue for many businesses today [6].

Churn prediction is a classification problem that differentiates between regular and churns customers. The traditional classifier looks for the minimal difference between classes. Many important uses were found for this technology. However, the statistics for predicting churn in the insurance industry are exceptional in that the regular database is much larger than the irregular database. Therefore, the default classifier is a dismal failure for our purpose. Predicting the actions of potential customers and identifying those who will not convert is part of De Caigny et al.'s [7] initial remit.

According to Levvel's studies, customers switch insurance companies for products in the property and casualty (P&C) sector, such as home, auto, and health insurance, only about once every ten years. Renter's insurance and term life insurance have much higher volatility, with consumers switching about once every five years as shown in Fig 1. Strategy: The insurance industry may want to diversify its products so that it can offer at least some products with lower churn [8].

The main contribution of this work is stated as follows,

- ✓ Customer churn prediction and early churner/non-churner identification is our main contribution.
- ✓ a novel AOA-SBLSTM-RNN model has been designed to proficiently forecast the occurrence of Customer Churn in the Insurance industry. Initially, the AOA model performs pre-processing to transform the original data into a useful format. In addition, the SBLSTM-RNN model is used to distinguish between churners and non-churners in the collected data.

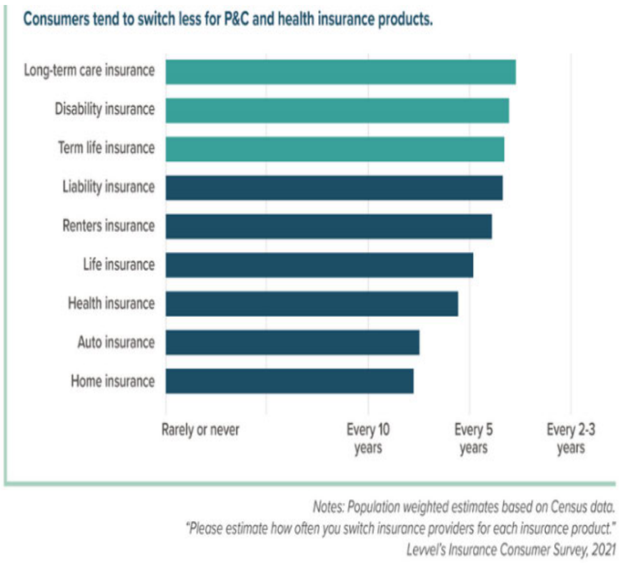

FIGURE 1. Consumers tend to switch to P&C and Health insurance products [8].

- ✓ To improve the CCP outcomes of the SBLSTM-RNN model, Improved Gravitational Search Optimization Algorithm (IGSA) based hyper parameter tuning process is involved in this study.
- ✓ Recent research findings are compared to the proposed investigation's prediction analysis for efficiency.

In this study, an AOA model is used to pick features for churn detection in the health insurance industry. To begin, the insurance company's dataset is pre-processed using the Z-score method. The standard dataset is divided (K-Fold) into a training set and a testing set. In phase1 AOA will reduce the dimensions to the get optimal feature subset. in phase II, the highest accuracy of churn prediction is done using the (SBLSTM-RNN) Using IGSA. Different metrics were used to conduct the analysis, and the results are represented in the Python environment. Topic-II defends the relevant literature, Topic-III details the methodology proposed, Topic-IV illustrates the findings, and Topic-V provides a conclusion.

II. LITERATURE REVIEW

Churn prediction has inspired the development of unique ways of using and adapting machine learning techniques in recent years. A variety of machine learning-related studies in the telecom industry, human resources, bank subscriptions, and financial services have been spurred by the high degree of interest in churn prediction [9]. Saradhi and Palshikar [10] looked at three machine learning approaches in the context of predicting employee attrition, which is analogous to predicting consumer churn [11]. A literature review and investigation comparing four models was proposed by Keramati et al. [12]. There have been other suggestions for comparative studies that use ensemble methods for machine learning. A list of frequently found models in churn analysis is provided by Umayaparvathi and Iyakutti [13] literature's survey on customer churn predictions in telecoms. The authors provide a list of four publicly available churn datasets as well as a brief discussion of the measures that could be used. The authors go into data collection, feature selection, model execution, and possible assessment techniques and metrics for churn model evaluation. Their survey closes with recommendations based on the literature. For churn, a variety of deep learning algorithms have been tested [14].

For comparison with logistic regression and artificial neural network concepts, Seymen et al. suggested a novel deep learning model. Their study offers a comprehensive evaluation of the literature on churn prediction using deep learning approaches. Outside of this sector, several evaluations dedicated to outlier discovery, which can be considered a dangerous case of customer turnover, have been offered [15]. The authors compare and contrast common shallow and innovative deep anomaly discovery algorithms. Pang et al. proposed an in-depth deep anomaly detection review that includes a classification system for deep learning approaches for outlier detection as well as a discussion of the challenges and future directions. [16]. N. Jajam et al. [17] proposed a daily churn forecasting model that forecasts churn using actions as a multivariate period On mobile telecom data, a statistical approach, an RFM model, an LSTM model, and a CNN concept were used. Everyday churn predictions beat monthly projections, they discovered. In churn forecasts, Umayaparvathi and Iyakutti [18] emphasize the necessity of feature selection. Deep learning approaches, they argued, were just as successful as traditional methods because they did not choose or extract characteristics from datasets like the others.

According to the study, the most credible data is obtained by combining support vector machines with a decision tree. The results demonstrate that the proposed churn prediction method is accurate. Customer churn detection was also improved using the stacking method [19]. To address CCP issues in the telecom industry, Amin et al. [20] studied data from a different company within the framework of JIT. The researchers put the proposed approach to the test using public data from two telecom companies. In the JIT-CCP context, cross-company samples can be used to evaluate the model's predictive ability, and it is clear that a diversified ensemble based on the JIT-CCP method is enhanced for single classifiers or a uniformly ensemble-based strategy. Scriney et al. [21] explained how to generate some of the missing CLV values. It has a complex database design and a unique method for filling in missing data pieces. An extensive knowledge discovery strategy using data from insurance policies was presented. The time-related aspects of the data set are compiled using the data collection TF/IDF descriptions. These extra qualities have been found in studies to improve accuracy, precision, and memory. Shirazi and Mohammadi [22] suggested a new way of forecasting churn. The purpose of the focused constructive retention strategy is to anticipate client churning. This proposed study would be useful to telecom companies to properly comprehend the danger of client turnover.

Börthas and Krange Sjölander [23] created a framework and applied two machine-learning algorithms, logistic regression and logit boost. In a test of the WEKA Machine Learning platform, real data from a U.S. company was used. There were several formats used to present the findings. Effectiveness was measured using the maximum profit uplift (MPU) metric, which looks at the massive profits that can be obtained using an uplift approach. Researchers have demonstrated using uplift models that show that retention efforts are more profitable since they exceed predicted models. De Caigny et al. [24] recently published a study that comprehensively investigated the use of convolutional neural networks in forecasting and demonstrated the benefits of CNN in processing text data. The authors partition continuous data variables using the CHAID technique and use the casual chart as the foundation of BBN to assess the CDR and other client services. Even yet, the link between the variables is not shown by this method [25]. Kirui .C et al. [26] employ ML techniques including MLP, SVM, and BN to estimate client churn. The method starts with Principal Component Analysis (PCA) data source pre-processing, followed by machine learning classification. Both MLP and BN are found to have worse rate prediction accuracy than SVM. Dalvi et al. [27] opted to implement machine learning methods again, this time focusing on DT and LR. This strategy is based on a mix of data mining techniques and comparative analysis. The results of the evaluation demonstrate that churn forecast accuracy has improved while the time required has decreased. However, only a few classes are included in the classification [28]. The RF classifier was used to predict worker turnover that is currently occurring owing to poor working conditions, low earnings, excessive job pressure, and low work satisfaction. They also used LDA and PCA to find the relevant characteristics to improve worker turnover prediction rates. This work would be more commendable if it produced any concrete results [29].

Staudemeyer and Omlin [47] used the KDD CUP 1999 dataset to assess the classification performance of a long short-term memory recurrent neural network (LSTM-RNN). Learning "memory" and building a model from time series data are both possible with LSTM networks. On their tweaked version of the KDD CUP 1999 dataset, the LSTM is trained and evaluated. The framework and parameters of the LSTM network were learned experimentally. The experimental findings were analyzed using a number of performance metrics. Their findings demonstrated that LSTM-RNN was able to learn all of the novel classes of attacks present in the dataset used for training. In addition, they discovered that ROC curves and area under the curve (AUC) were suitable for assessing LSTM-RNN.

The goal of this paper is to define the process followed in developing a stacked bidirectional long short-term memory (SBLSTM) and RNN model for CCP using the arithmetic optimization algorithm (AOA). The proposed

AOA-SBLSTM-RNN model intends to proficiently forecast the occurrence of CC in the Insurance industry. Initially, the AOA model performs pre-processing to transform the original data into a useful format. In addition, data is separated into churners and non-churners using the SBLSTM-RNN approach. To improve the CCP outcomes of the SBLSTM model, an optimal Hyperparameters tuning process using IGSA is developed. A widespread simulation analysis of the AOA-SBLSTM-RNN model is tested using a benchmark Dataset which includes 42213 records. There are 33908 observations in the training set, and 11303 in the testing set. In the insurance Churn Prediction-Machine Hackthan dataset, the desired variable is The class labels (factor:16).

III. PROPOSED CHURN PREDICTION MODEL

In this paper, The Authors propose using an Ensemble Deep Learning based on the classification model and feature extraction with AOA for handling insurance client data. Figure 2 is a representation of the model, and in this article, we employed three insurance datasets to assess the effectiveness of different settings. To better predict the incidence of Customer Churn in the Insurance industry, a novel AOA-SBLSTM-RNN model has been developed in this study. The AOA model first processes raw data into a more usable form through a process called pre-processing. In addition, data is separated into churners and non-churners using the SBLSTM-RNN model. To improve the CCP outcomes of the IGSA model used in an optimal Hyperparameters tuning process. Fig. 2 demonstrates the overall process of the AOA-SBLSTM-RNN-IGSA technique.

Standard pre-processing steps are performed on the collected insurance data, including the elimination of missing values, the conversion of text to numbers, and the normalization of the data. Following that, utilizing several Deep learning models (LSTMs, RNNs, RBMs, RBFNs, and LSTM-RRN) are used along with feature extraction with AOA. characteristics for such insurance client churn prediction datasets are discovered. This finding proves that integrating

SBLSTM-RNN with the IGSA approach yields the best performance gains and accuracy.

This method for predicting customer churn employs a 4-stage setup.

1. The effectiveness of the base classification algorithms when dimension reduction is not applied.

2. The Effectiveness of Base Classification Algorithms Taking Dimension Reduction into Account

3. The effectiveness of deep learning classification algorithms when dimension reduction is not applied

4. The effectiveness of deep learning classification methods when dimension reduction is applied

Finally, the proposed method is evaluated using a number of different classification algorithms that already exist.In this paper, 3 Datasets were used in Insurance Churn Prediction,

A. INSURANCE CHURN PREDICTION -MACHINE HACKTHAN

The dataset that was used contains a total of 42213 records. There are 33908 observations found in the training section, and there are 11303 observations found in the testing section. The target variable in the insurance market is the Class labels (factor:16). Churn the dataset for the Hackthan prediction machine. [41]

B. HEALTH-INSURANCE-CROSS-SELL-PREDICTION

The creation of a model that can predict whether or not a customer is interested in purchasing insurance is beneficial to the company because it enables the company to modify its marketing strategy to target customers who are interested in purchasing insurance. As a result, the company is able to expand its customer base and revenue. [42]

C. HEALTH INSURANCE LEAD PREDICTION

The creation of a method to predict whether a client is interested in insurance is extremely beneficial to the company because it enables the company to tailor its promotional plan to reach out to those clients and, as a consequence, increase both its business model and its profit. You now have access to data on attributes (such as Gender, Age, and Region code type), policies (such as premium and sourcing channel), and so on, which you can use to determine whether or not the customer has an interest in Insurance. [43]

D. PROBLEM STATEMENT

Insurance companies competed intensely for businesses all around the globe. Due to the obvious recent increase in the highly competitive environment of health insurance businesses, clients are migrating. It's unclear if there is evidence of switching behavior and which customers go to a competitor. When there are so many different pieces of information recorded from millions of clients, it's tough to study and comprehend the causes of a customer's decision to switch insurance carriers. In an industry where customer retention is equally important, with the earlier being the costlier method, insurance companies depend on the information to evaluate client behavior to minimize loss. Insurance firms may design techniques to prevent customers from really transferring if they know whether or not they are likely to do so ahead of time. The ability to accurately estimate future attrition rates is crucial since it helps the organization comprehend future earnings. Churn level forecasts may also help your firm find and boost areas where customer service is less. As a result, Our AOA-SBLSTM-RNN-IGSA Proposed Model will be used to investigate machine learning-based customer churn detection for insurance data.

E. DATA PROCESSING

To begin, the raw data is transformed into a more usable format by applying a z-score normalization-based preprocessing in the AOA-SBLSTM model. In calculating a

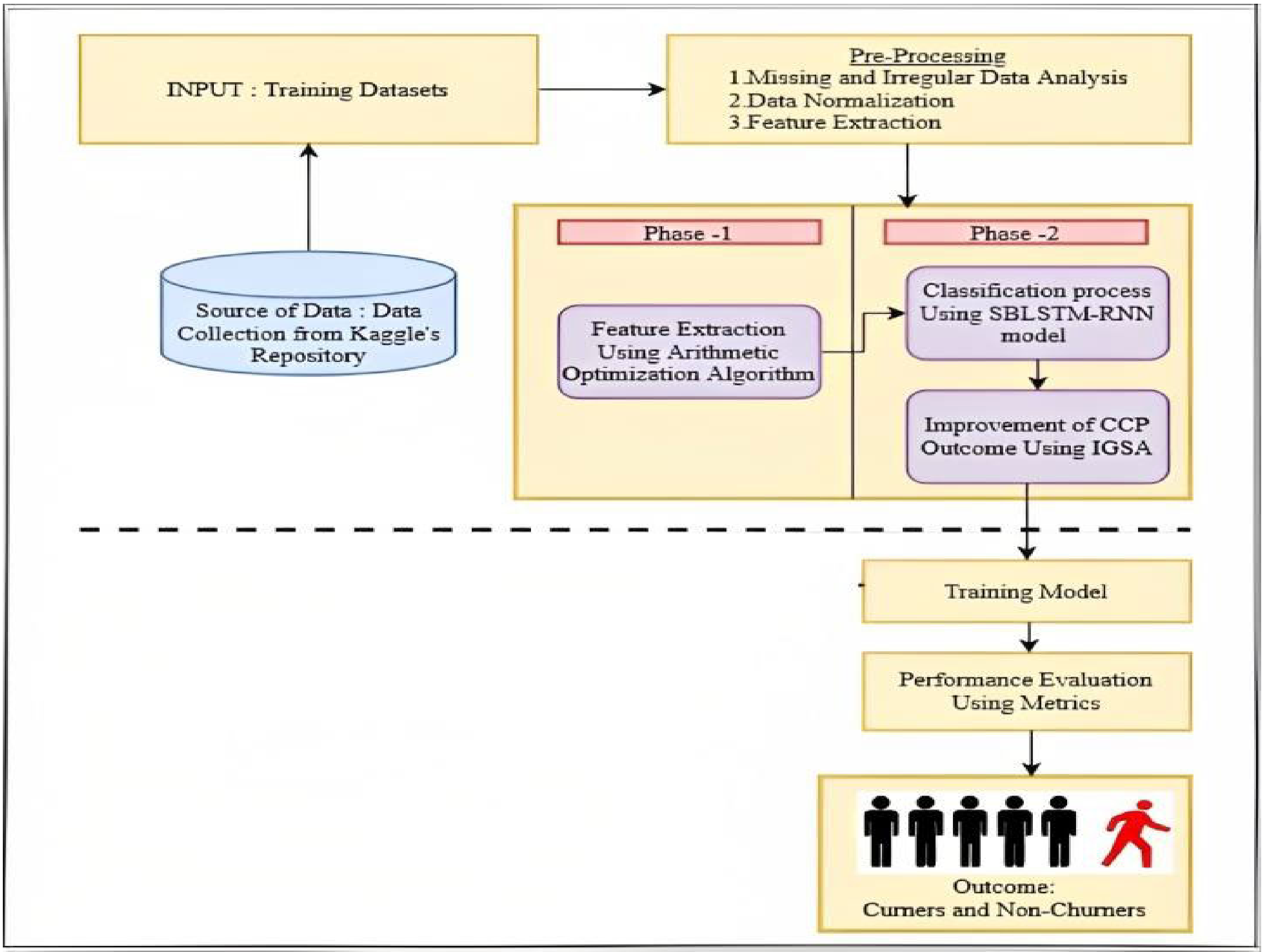

FIGURE 2. The overall process of the AOA-SBLSTM-RNN-IGSA Proposed Model.

Z-score, the standard deviation of the original data is typically used to normalize the final score. As might be expected, this normalization approach fares well when the matcher's average score and the range of possible scores are known in advance. Equation 1 shows the normalized score.

$$New\ value = \frac{(x-\mu)}{\sigma} \tag{1}$$

F. K-FOLD CROSS-VALIDATION

A notable cross-validation approach is the K-fold method. All data is randomly and evenly spread into k folds. The test data is selected, while the remaining sets are used as training data. This is done until each set of data has been used as test data, indicating that the testing has been run k times. The power of the suggested technique was evaluated using the k-fold cross-validation approach in the studies, with k set to 10.

G. DATA SPLITTING

In fact, data splitting is a technique for segmenting data sets into training (or calibration) and testing (or prediction) sets. This step is typically carried out after the spectra of the samples have been corrected for distortion and unwanted variation during the pre-processing phase. The training dataset would contain all the data used to train the model. The testing dataset contains the information used to test the trained and validated strategy. It shows how accurate our overall model is and how often we would make a wrong prediction. Our research indicates that there are a total of 33,908 observations in the training set and 11,303 in the testing set.

H. ARITHMETIC OPTIMIZATION ALGORITHM (AOA) APPROACH IN FEATURE EXTRACTION

Abualigah and colleagues [16] introduced a new meta-heuristic model called AOA in 2021. The idea behind this method is to perform operations in mathematics that involve more than two of the four basic arithmetic operators (Division, Multiplication, Addition, and Subtraction). AOA is a simple framework with low computation complexity, and it is related to the sine-cosine algorithm (SCA). Assuming that M&D companies are turning out massive stages from the iterations, the exploration phase is where the majority of the work is done. Equation 2 summarizes the formulation as follows.

$$X_i(t+1) = \begin{cases} Xb(t)/(MOP + eps).((UB - LB)\mu \\ \qquad +LB, rand < 0.5 \\ Xb(t).MOP.((UB - LB)\mu + LB), rand \geq 0.5 \end{cases} \tag{2}$$

the very simple positive number denoted by eps, and the constant co-efficient denoted by l (0.499) are both factors in the method's design. Formulationally, MOP involves a non-linear decrease from one to zero as the iterations progress, as shown in Equation 3 below.

$$\mathrm{MOP} = 1 - (\frac{\mathrm{t}}{\mathrm{T}})^{1/\alpha} \tag{3}$$

where an is the constant value set at 5 in accordance with the AOA. It is possible to see that both the M and D operators in Eq. (3) produce highly random starting points for their optimum-searching agents. On the flip side, S and A operators were implemented to place more emphasis on local exploitation, thereby reducing the number of stages beneath the searching space. Equation 2 gives us the formula in mathematics:

The right equilibrium between discovery and use is crucial for reaching peak performance in any method. Equation 5 describes how the AOA parameter MOA was used to toggle between exploration and exploitation over the course of iterations.

$$\mathrm{MOA}\ (\mathrm{t}) = \mathrm{Min} + \mathrm{t}\left(\frac{\mathrm{Max}-\mathrm{Min}}{\mathrm{T}}\right) \tag{4}$$

whereas Min and Max are constant values. Based on Eq. (4), 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 pseudocode of AOA is demonstrated in Algorithm 1.

To improve classification precision, the AOA method derives a fitness function. It specifies a positive integer to describe the practical precision of the potential answers. The classification error rate is considered the fitness function in our work, and its minimization is the goal. In Equation 6, we see that the error rate is minimal for the best solution and increases for the worst. Figure 3 shows the overall flow of AOA.

$$\begin{aligned} \text{fitness}\ (\mathrm{X_i}) &= \text{ClassifierErrorRate}\ (\mathrm{X_i}) \\ &= \frac{\textit{number of misclassified samples}}{\textit{Total number of samples}} * 100 \end{aligned} \tag{5}$$

Algorithm 1 Pseudo-code of AOA

Initializing the parameters population size (N) and maximal iterations (T)
Initializing the places of every searching agent Xi (i=l,2,3,..., N)
Set the parameters a. ji, Min, and Max
while (t<=T)
Compute the fitness of every search agent Upgrade Optimum Fitness, Xb Compute the MOP by Eq. (3)
Compute the MOA by Eq. (5)
For all search agents
If rand > MOA
Upgrade place by Eq. (4)|
Else
Upgrade place by Eq. (4)
End if
End for
t=t+l
End While
Return optimum Fitness, Xt>.

I. CUSTOMER CHURN ANALYSIS USING LSTM-RNN MODEL

The LSTM-RNN model is applied for the churn classification procedure. The aim is to discover a DL framework including an attention layer to understand churn classification accuracy from the input dataset. Hence, there are a few more processing stages for the proposed LSTM-RNN model. First is the convolution of features. At that stage, the input data is fed into the LSTM-RNN architecture [20]. This phase aims to extract higher-level semantic features from the series of words. Also, the LSTM-RNN architecture discovers the temporal relationships amongst features and generates a feature vector.

In addition, assume the semantic meaning of input data and generate a second set of labels with value allocated to the churn expressed in the input data to train the datasets: Churn=1 and non-churn=0. Assuming that these assignments of churn labels are subjective, thereby generating a matrix layer that is fed to LSTM-RNN towards accomplishing SA results in Fig 4.

LSTM cells have some drawbacks when compared to simple RNN cells. They are more computationally expensive and require more memory and time to train and run due to their additional parameters and operations. RNNs are superior to other types of neural networks because their architecture, which is based on deep neural networks, allows them to process information in both sequential and parallel fashion. Adding memory cells to a neural network allows it to mimic the brain's processing abilities.

The LSTM-RNN variables are: 1. There are 41 input layers, 82 iterations in the hidden layers, 5 classes, 1000 epochs of training, a learning rate of 0.001, and an RMSProp the optimization algorithm.

The main Disadvantages of LSTMs are:

- ✓ The training of LSTMs takes more time.
- ✓ LSTMs require a greater amount of memory in order to be trained.
- ✓ It is simple to overfit an LSTM model.

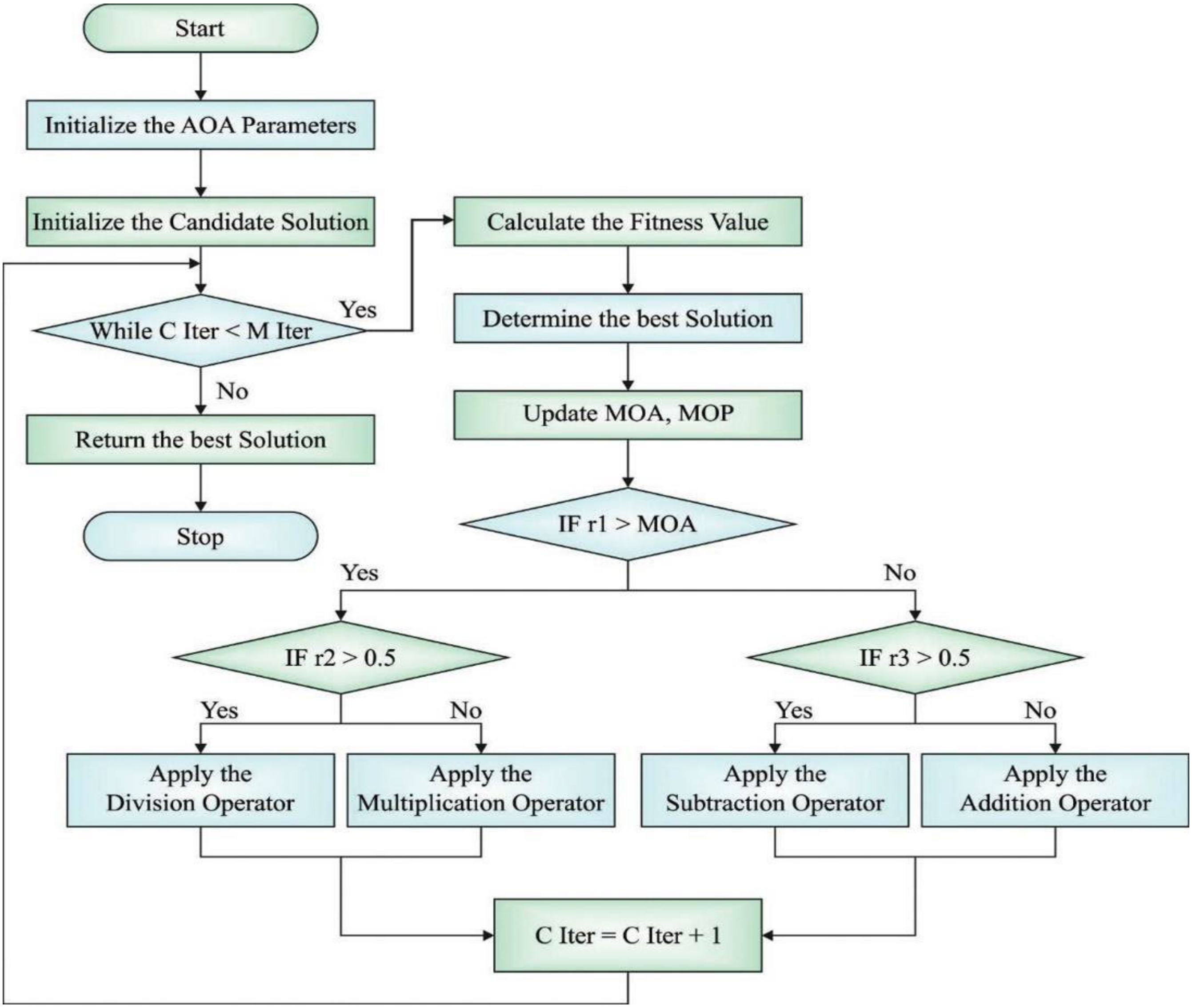

FIGURE 3. Flowchart of AOA.

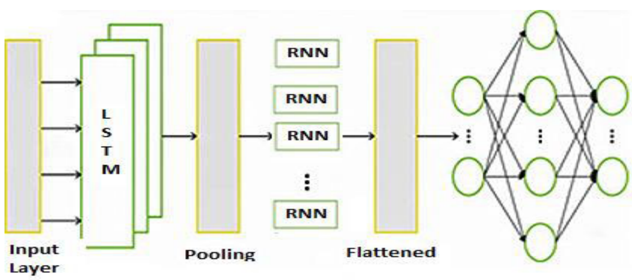

FIGURE 4. LSTM-RNN structure.

- ✓ Dropout is significantly more difficult to implement in LSTMs.
- ✓ LSTMs are extremely sensitive to the random weight initializations that are used.

The following are the primary drawbacks of RNNs:

- ✓ Training RNNs.
- ✓ The problem of the gradient either vanishing or exploding.
- ✓ RNNs cannot be layered one on top of another.

The Draw backs of LSTM-RNN are:

- ✓ Methods of instruction that are laborious and time-consuming.
- ✓ Difficulty in processing sequences that are longer

J. SBLSTM-RNN BASED PREDICTION

There are two primary reasons to use an ensemble rather than a single model, and they are related; these reasons are as follows: Advantages of Ensemble Deep Learning SBLSTM-RNN Prediction Model

1) PERFORMANCE

An ensemble of models can make more accurate predictions and achieve higher levels of performance than any individual model that it contributes.

2) ROBUSTNESS

An ensemble helps reduce the variation in the predictions and the overall performance of the framework.

At this point in the process, the SBLSTM-RNN model is utilised to classify the data as either churners or non-churners. The normal structure of SBLSTM is able to resolve the gradient vanishing problems in an effective manner and efficiently transport valuable data throughout the SBLSTM network [15]. The Recurrent Neural Network (RNN) was unable to accurately capture the long-term dependency that existed between feature vectors. The LSTM cell comprises of input gate (it), forget gate (fi), and output gate (0t). That gate controls a memory cell activation vector.

3) FORGET GATE

Using the current input xt and the hidden layer ht_1, the "forget gate" can determine how much information from the previous layer, ct_1, should be forgotten and how much should be retained. The formula for the forget gate is given in Equation 6.

$$\mathrm{f_t} = \mathrm{Sigmoid}(\mathrm{w_{xf}x_t} + \mathrm{w_{hf}h_{t-1}} + \mathrm{b_f}) \tag{6}$$

whereas $_{\mathrm{bf}}$, $\mathrm{w_{xf}}$, and $_{\mathrm{wh}}$f denotes the bias vecto; the weighted matrix is among xt and f; and the weighted matrix is among $_{\mathrm{ht_}}$1 and $_{\mathrm{ft}}$z, correspondingl

4) INPUT GATE

It can be used in Equation 7 to find out how many of the network's input xt should be kept in the current cell state ct.

$$\mathrm{i_t} = \mathrm{Sigmoid}(\mathrm{w_{xi}x_t} + \mathrm{w_{hi}h_{t-1}} + \mathrm{b_i}) \tag{7}$$

In which $\mathrm{b_{f@comm}}$ $\mathrm{a_{wx}}$i, and $_{\mathrm{wh}}$i denotes bias vector, the weighted matrix is among xi and it, and the weighted matrix is among $_{\mathrm{ht_}}$1 and it, correspondingly.

5) OUTPUT GATE

The current output ht can be used to calculate the total number of inputs to the LSTM in the cell state ct. The LSTM gate is a fully connected network with a vector input and a real output. Equation 8 can be used to describe the output gate.

$$\mathrm{o_t} = \mathrm{Sigmoid}\,(\mathrm{w_{x0}x_t} + \mathrm{w_{ho}h_{t-1}} + \mathrm{b_o}) \tag{8}$$

where as bf, $_{\mathrm{wx}}$o, and $_{\mathrm{who}}$denotes the bias vectors, the weight matrix is among xt and 0; also the weight matrix is among $\mathrm{h_{t\_1}}$ and 0t, correspondingly

The LSTM cell's final output consists of the cell output state ct and the layer output ht, which are defined in Equation 9,10, respectively.

$$\mathrm{C_t} = \mathrm{f_t} \odot \mathrm{C_{t-1}} + \mathrm{i_t} \odot \tilde{C}_t \tag{9}$$

$$\mathrm{h_t} = \mathrm{o_t} \odot tan\mathrm{h}(\mathrm{C_t}) \tag{10}$$

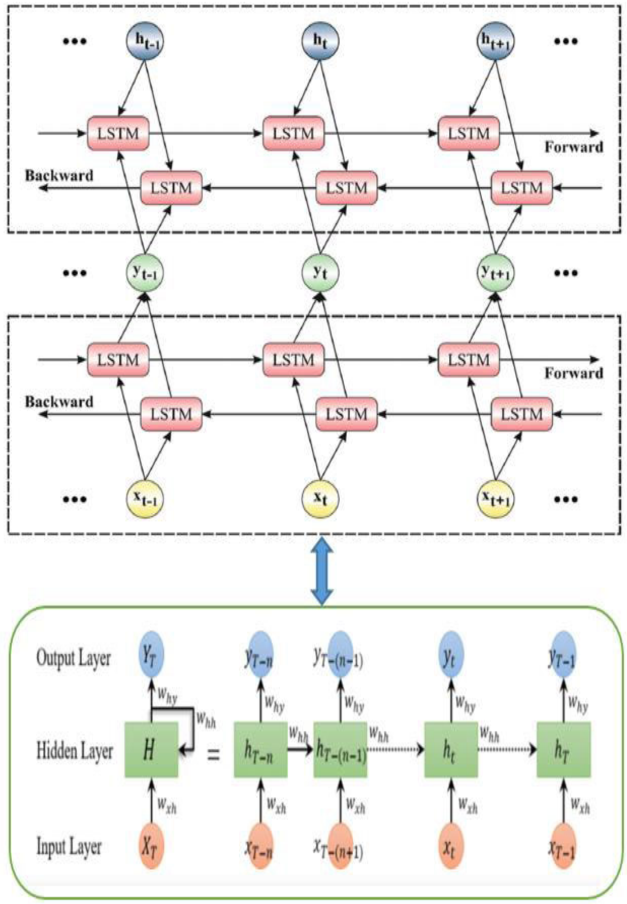

FIGURE 5. Structure of SBLSTM-RNN Prediction Model.

The intermediate cell input is denoted as ct that is formulated in Equation 11:

$$\tilde{C}_t = tan\mathrm{h}(\mathrm{W_{xc}x_t} + \mathrm{W_{hc}h_{t-1}} + \mathrm{b_c}) \tag{11}$$

here, the bias vector is represented by bf, $_{\mathrm{wx}}$c, and $_{\mathrm{wh}}$; the weight matrix is between vi and ct; and the weight matrix between $\mathrm{h_{t_1}}$zand $_{\mathrm{ct}}$. Stacked bidirectional long short-term memory (BilSTM) [15] is depicted in Fig. 5.

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 Stacked 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 and backward variables in the stacked BiLSTM are autonomous of each othe's, even though word embedding is distributed. At last, the hidden layer of stacked BiLSTM is determined that concatenate the vector of the forward and backward directions during time step t in Equation 12

$$\mathrm{h_t} = \left[\vec{h}_t, \overleftarrow{h}_t\right] \tag{12}$$

TABLE 1. Proposed stacked BiLSTM(SBLSTM)-RNN model description.

Layer	Number of Nodes/ Percentage Rate	Output Shape	Number of Parameters Received	Activation function Used
Bidirectional LSTM layer	256	(None, 30, 512)	528384	Sigmoid
Dropout Layer	20%	(None, 30, 512)	0	None
Bidirectional LSTM layer	128	(None, 30, 256)	656384	Sigmoid
Dropout Layer	20%	(None, 30, 256)	0	None
Bidirectional LSTM layer	64	(None, 30, 128)	164352	Sigmoid
Dropout Layer	20%	(None, 30, 128)	0	None
Bidirectional LSTM layer	16	(None, 32)	18560	Sigmoid
Dropout Layer	20%	(None, 32)	0	None
Dense layer	8	(None, 8)	264	None
Dense layer	4	(None, 4)	36	None
Dense layer	2	(None, 2)	10	None
Dense layer	1	(None, 1)	1	Sigmoid

Further, the Ensemble deep learning method improves the accuracy of the classification. But the stacked BiLSTM-RNN is very effective in the shallow learning method. Hence the presented method determines a stacking BiLSTM-RNN to employ the local context and latent symmetry complex data. In a stacked BiLSTM, the output of the lower layer is used as the input for the RNN that is in the higher layer.

The SBLSTM-RNN model is used in the proposed method to make such forecasts. In this paper, we apply an SBLSTM-BRNN model that comprises a stack of four Bidirectional LSTM layers and four dense layers. Each layer of this SB LSTM-RNN model consists of 256, 128, 64, and 16 nodes. Following each of these is a 20% dropout regularization. Then, four stacked dense layers with 8, 4, 2, and 1 nodes each are created. To activate the first four SBLSTM layers and the last dense layers, a sigmoid activation function is used. After that, an Adam optimizer and a binary cross entropy loss function are used to create a compilation of the aforementioned layers. This model's development requires 100 epochs and 32-person batches. Adjusting the hyper-parameters helps find the optimal solution to the problem. After this model is built, it is fed training data. The parameters of this neural network model can be trained to produce a prediction in one of 1,367,993 different ways. Table 1 summarizes information about the number of layers, layer types, activation functions, output shapes, and accepted parameters for each layer.

K. HYPER PARAMETER TUNING USING IMPROVED GRAVITATIONAL SEARCH OPTIMIZATION ALGORITHM (IGSA)

Finally, to boost the classification efficiency of the SBLSTM-RNN model, the IGSA-based hyper parameter tuning process is involved in this study. The GSA that is inspired by the law of mass and gravity interactions. At t^th iteration, the force acts on particle x_j (t) from particle x_(j(t)) can be described by Equation 13,

$$F_{i,j}^{d}(t) = G(t) * \frac{I_{pi} * I_{aj}}{R_{i,j}(t) + \varepsilon} * \left(x_{j,d}^{t} - x_{i,d}^{t}\right) \tag{13}$$

In Eq. (13), I_{pi}indicates the active gravitational mass associated with $\chi_{j(t),I_{aj}}$ particle is the passive gravitational mass associated with $x_i(t)$particle, ε denotes a smaller constant, G(t) shows the gravitational constant at t iteration as follows in Equation 14:

$$G(t) = G_0 * e^{-\alpha * \frac{t}{N}} \tag{14}$$

where $G_0 = 100$ and $\alpha = 20$. $R_{i,j}(t)$ denote the Euclidian distance between 2 particles x(t) and xj(t), defined below Equation 15.

$$R_{i,j}(t) = \sqrt{\sum\nolimits_{k--1}^{D} (x_{i,k}^{t} - x_{j,k}^{t})^2} \tag{15}$$

At t^{th}iteration, the overall force acts on the x(t) particle in d dimension is determined as Equation 16,

$$F_{i}^{d}(t) = \sum\nolimits_{j=1, j\neq i}^{M} rand_j * F_{i,j}^{d}(t) \tag{16}$$

In Eq. (36), $rand_j$denotes a random integer within [1, 0], and the acceleration of particle X(t) in the d dimension is described by Equation 18:

$$a_i^d(t) = \frac{F_i^d(t)}{I_{ii}(t)} \tag{17}$$

In Eq. (17), I(t) indicates the inertial mass of particle $x_i(t)$one particle upgrades its position and velocity based on the acceleration, as follows in Equation 18 and 19.

$$v_{i,d}(t+1) = rand_i * v_{i,d}(t) + a_i^d(t) \tag{18}$$

$$x_{i,d}(t+1) = x_{i,d}(t) + v_{i,d}(t+1) \tag{19}$$

The GSA approach upgrades the inertial and gravitational masses using the subsequent Equation 20 and 21:

$$I_{ai} = I_{pi} = I_{ii} = I_{i'} i = 1, 2, \ldots, M$$

$$i_i(t) = \frac{f_i(t) - worst(t)}{best(t) - worst(t)} \tag{20}$$

$$I_i(t) = \frac{i_i(t)}{\Sigma_{j=1}^{M} i_j(t)} \tag{21}$$

In Eq. (21), $f_i(t)$ denotes the fitness value of particle $x_i(t)$, best(t) and worst (t) characterize both fitness values amongst every particle, correspondingly, and it is determined by the succeeding expression in Equation 22,22.

$$best$$

$$best(t) = j \in \{1, 2, \ldots, M\}^{f_j(t)} \tag{22}$$

$$worst$$

$$worst(t) = j \in \{1, 2, \ldots, M\}^{f_j(t)} \tag{23}$$

The IGSA is developed in this research by fusing GSA with DOL (dynamic oppositional-based learning). An alternative to the current solution is developed using the OBL methodology. In an effort to increase the rate of convergence, it seeks

Actual	Predicted	
	Churn	*NonChurn*
Churn	A11	A12
NonChurn	A21	A22

FIGURE 6. Confusion Matrix model.

to define the best possible solution. The opposite (X^0) of real number (X∈[U,L]) is evaluated by the Equation 22 and Equation 24.

$$X^0 = U + L - X \tag{24}$$

Opposite point: Assume that $X = [X_1, X_2, \ldots, X_{Dim}]$ refers to the point in the *Dim*-dimension search domain, and $X_1, X_2, \ldots, X_{Dim} \in R$ and $X_j[U_j, L_j]$. Therefore, the opposite point (X^0) of X is formulated in the following Equation 25:

$$X_j^0 = UB_j + L_j - X_j, \, where \, j = 1 \ldots .D. \tag{25}$$

In addition, the values of the fitness function are used to select the two most useful points (X^0 and X), while the other is ignored, if $(X) \leq f(X^0)$, then (X) is maintained for the minimization problem, and vice versa. In relation to the inverse, Equation 26 represents the dynamic opposite preference (X^{DO})) of the value X:

$$X^{Do} = X + w \times r_8 \left(r_9 \times X^0 - X \right), w > 0 \tag{26}$$

where r_8 and r_9 denote random integers within [01]. w denotes weighting agent. Subsequently, the dynamic opposite value (X_j^{DO}) of X is equivalent to $[X_1, X_2, \ldots, X_{Dim}]$, which is shown below Equation 27:

$$X_j^{Do} = X_j + w \times rand \left(rand \times X_{j^0} - X_j \right), w > 0 \tag{27}$$

Consequently, DOL optimization initiates by generating the initial solution ($X=(X_1, \ldots, X_{Dim})$ and evaluating the dynamic opposite value (X^{Do}) based on the above formula. Then, according to the fitness value, the optimum solution from $X^{Do} and X$ is employed, and another one is omitted.

Algorithm 2 Pseudocode of IGSA

Input: the parameters $\boldsymbol{M}, \boldsymbol{N}, \delta, G_0, \alpha$.
Begin
Step 1: initialize $\boldsymbol{M}$ particles x,(t) arbitrarily $0 < \mathrm{i} < \boldsymbol{M}$, iterative times t = 1;
Step 2: calculate (i), $\boldsymbol{0} < \mathbf{i} < \boldsymbol{M}$ upgrade $p_g(t)$, if it fulfills (t>N or precision $< \delta$ then return to step Step4; or else, proceed to step 3;
Step 3: upgrade $\boldsymbol{G}$(t), upgrade best (t) and worst (t) based on Eq. (23), upgrade I(t)based on Eq. (24), calculate F_i^d (t) based on Eq. (17), calculate α_d^i (t) based on Eq, (18),
upgrade $v_{i,d}(t+1)$ and $X_{i,d}(t+1)$ based on Eq. (37), iterative times t = $t+1$; proceed to step S2:
Step 4: output the better outcomes.
End

TABLE 2. Dataset1: Insurance churn prediction -machine hackthan.

Classifier	SVM	DT	RF	NB	KNN
Accuracy	83.69	82.81	85.21	86.27	**87.13**
True Churn	89.58	83.78	86.52	95.35	95.89
False Churn	12.77	36.96	25.1	11.21	15.69
Specificity	87.23	63.04	74.9	88.22	84.31
Precision	90.42	81.29	89.53	89.86	90.23

IV. EXPERIMENTS AND RESULTS

A. PERFORMANCE METRICS

The confusion matrix plays a crucial role in comparing and analyzing classifier performance. Figure 6 shows the number of incorrect predictions made by the classifier in columns A12 and A21, as well as the number of instances where the classifier made a positive prediction in column A11. Equations (28)-(32) depict the following: See below for illustrations of accuracy, true churn, false churn, specificity, and precision.

$$\text{Accuracy} = (A11 + A22)/(A11 + A12 + A21 + A22) \tag{28}$$

$$\text{True churn} = A12/(A11 + A12) \tag{29}$$

$$\text{False churn} = A21/(A21 + A22) \tag{30}$$

$$\text{Specificity} = A22/(A21 + A22) \tag{31}$$

$$\text{Precision} = A11/(A11 + A21) \tag{32}$$

DT, SVM, KNN, NB, and RF are some of the five base classification models used in this experiment, while Deep Learning classification techniques like LSTMs, RNNs, RBMs, RBFNs, and SBLSTM-RNNs are also employed. Method of reducing dimensions in this experimental setup, AOA is used to extract useful features from the datasets. Finally, several existing datasets are used to evaluate the effectiveness of the proposed model strategies.

B. THE PERFORMANCE OF BASE CLASSIFICATION ALGORITHMS WITHOUT DIMENSION REDUCTION

The presented CCP results will be compared to those of existing techniques such as DT, KNN, SVM, NB, and RF in the experimental studies. Classification is a technique that groups together and represents similar information using class labels. The first step in classification is to divide the entire set of data into testing and training partitions. The training datasets are used to structure the classification rules, and the testing dataset is used to run the test.

Training and testing of churn predict dataset (calibration) are taken from Kaggle's repository for this task, along with two other datasets. To predict whether or not an insurance policyholder will cancel their coverage based on these factors, a Learning model must be developed. The efficiency of the Base classification models with tested parameters is shown in Tables 2, 3, and 4. When compared to classifications performed without the pre-processing, the results of

TABLE 3. Dataset2: Health-insurance-cross-sell-prediction.

Classifier	SVM	DT	RF	NB	KNN
Accuracy	86.49	85.98	88.96	81.25	86.21
True Churn	92.68	92.45	95.87	89.13	95.02
False Churn	11.64	13.68	14.02	16.33	25.93
Specificity	88.36	86.32	85.98	84.12	74.07
Precision	92.68	92.29	90.18	89.78	93.03

TABLE 4. Dataset3: Health insurance lead prediction.

Classifier	SVM	DT	RF	NB	KNN
Accuracy	88.47	84.36	90.57	87.22	88.23
True Churn	97.06	91.43	93.82	93.39	94.69
False Churn	3.88	14.17	10.14	15.55	13.04
Specificity	96.12	85.83	89.86	84.45	86.96
Precision	90.86	91.36	94.87	92.77	95.35

TABLE 5. Dataset1: Insurance churn prediction -machine hackthan.

Classifier	SVM+ AOA	DT+ AOA	RF+ AOA	NB+ AOA	KNN+ AOA
Accuracy	81.69	**90.81**	89.21	86.27	83.13
True Churn	91.06	97.43	93.82	93.39	94.69
False Churn	3.88	14.17	10.14	15.55	13.04
Specificity	87.23	63.04	74.9	88.22	84.31
Precision	89.42	84.29	88.13	89.86	91.23

TABLE 6. Dataset2: Health-insurance-cross-sell-prediction.

Classifier	SVM+ AOA	DT+ AOA	RF+ AOA	NB+ AOA	KNN+ AOA
Accuracy	86.49	**91.98**	88.96	81.25	86.21
True Churn	92.68	92.45	95.87	89.13	95.02
False Churn	11.64	13.68	14.02	16.33	25.93
Specificity	88.36	86.32	85.98	84.12	74.07
Precision	92.68	92.29	90.18	89.78	93.03

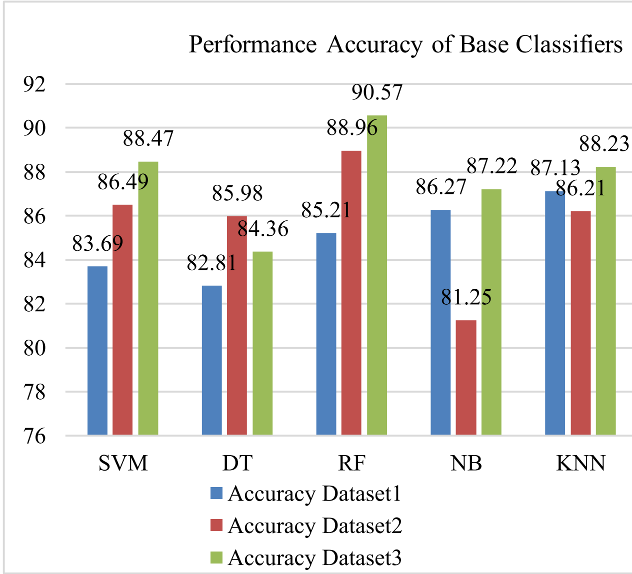

FIGURE 7. Performance Accuracy of Base Classifiers.

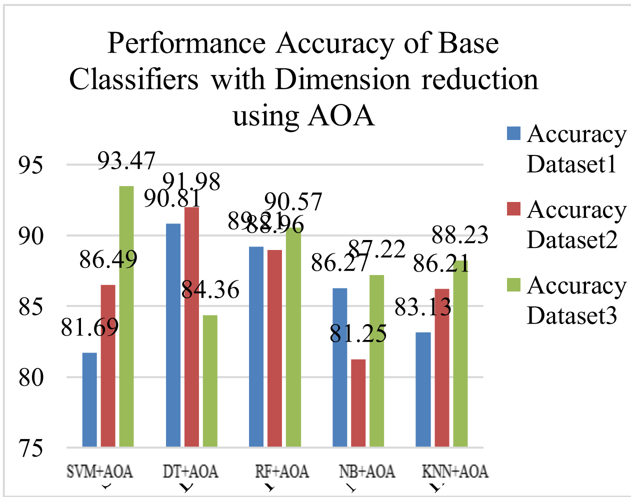

FIGURE 8. Performance Accuracy of Base Classifiers with Dimension reduction using AOA.

classifications performed using data processed in this model are satisfactory. KNN has the highest True churn (95.89) and Accuracy (87.13) in dataset:1, making it the most successful classifier overall and meeting the majority of the highlighted, most important goals. For datasets 2 and 3 Random Forest (RF) classifier shows the highest accuracy 88.96 and 90.57 respectively and which are underlined with bold letters. The accuracy values are shown in the below Figure 7 for better understand ability.

C. PERFORMANCE OF BASE CLASSIFICATION ALGORITHMS WITH DIMENSION REDUCTION (AOA)

Tables 4, 5 and 6 show the effectiveness of the Base classification models with Dimensionality reduction in terms of AOA used, and various parameters were tested. Data pre-processed in this model performs better enough for the entire classification methods in comparison to a classification performed without the pre-processing. Among classifiers examined, DT+AOA shows the best Accuracy of 90.81,91.98 in dataset1,2 respectively, and the greatest True churn of 94.69 and 95.02, SVM+AOA shows 93.47 accuracies in dataset 3, all highest accuracy values are underlined in bold letters. The accuracy comparison values are shown in the below Figure 8 for better understandability.

D. PERFORMANCE OF DEEP LEARNING CLASSIFICATION ALGORITHMS WITHOUT DIMENSION REDUCTION

Deep learning models have been utilized in computer vision, pattern matching, and NLP in the past. Deep learning methods can be used to learn multilevel representations. Deep

TABLE 7. Dataset3: Health insurance lead prediction.

Classifier	SVM+ Hybrid PCA	DT+ Hybrid PCA	RF+ Hybrid PCA	NB+ Hybrid PCA	KNN+ Hybrid PCA
Accuracy	93.47	84.36	90.57	87.22	88.23
True Churn	92.68	92.45	95.87	89.13	95.02
False Churn	11.64	13.68	14.02	16.33	25.93
Specificity	96.12	85.83	89.86	84.45	86.96
Precision	89.86	81.36	95.87	93.76	94.34

TABLE 8. Dataset1: Insurance churn prediction-machine hackthan.

Classifier	LSTM	RNNs	RBMs	RBFNs	SBLSTM-RNN
Accuracy	**95.47**	90.36	91.57	95.22	90.23
True Churn	90.33	91.23	94.22	90.23	94.33
False Churn	10.64	14.68	13.02	13.33	15.53
Specificity	95.34	84.44	90.86	86.33	85.2
Precision	86.65	83.33	95.87	95.32	93.94

TABLE 9. Dataset2: Health-insurance-cross-sell-prediction.

Classifier	LSTM	RNNs	RBMs	RBFNs	SBLSTM-RNN
Accuracy	92.47	90.36	91.57	93.22	**96.22**
True Churn	91.32	90.12	93.11	89.23	93.22
False Churn	11.63	13.67	12.034	13.33	9.53
Specificity	94.34	83.44	91.86	87.33	82.23
Precision	87.55	84.33	94.87	96.32	94.22

TABLE 10. Dataset3: Health insurance lead prediction.

Classifier	LSTM	RNNs	RBMs	RBFNs	SBLSTM-RNN
Accuracy	93.65	91.63	91.23	94.27	**96.81**
True Churn	90.23	91.11	92.98	90.32	94.11
False Churn	10.39	12.67	11.44	10.75	9.12
Specificity	93.35	84.45	92.65	88.22	81.67
Precision	88.44	83.21	95.22	95.88	93.34

learning is a subfield of machine learning. However, as its importance grows, it has evolved into a distinct academic field. In this paper five types of deep learning models are used to find the classification Accuracy. The models are LSTM, RNN, SBLSTM-RNN, RBM and RBNF. SBLSTM-RNN with IGSA performs well as compared with other deep learning models.

Table 8, 9, and 10 shows the effectiveness of the Deep classification models without Dimensionality reduction used,

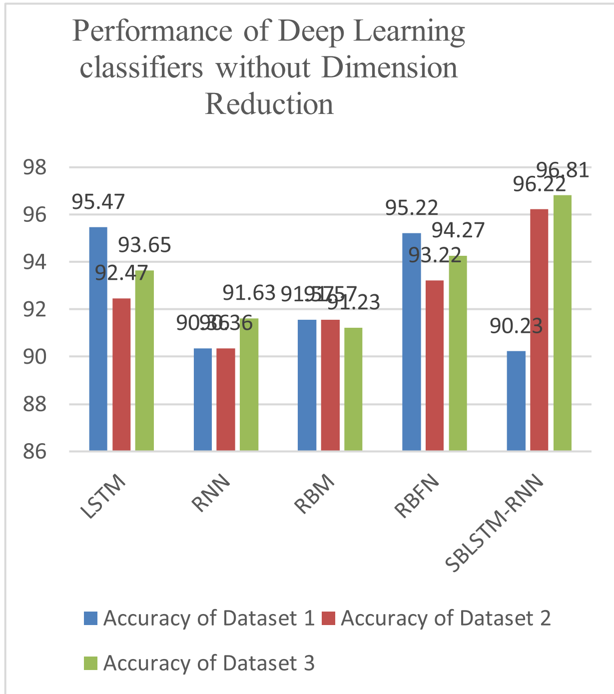

FIGURE 9. Performance of Deep Learning Classifiers without Dimension reduction.

TABLE 11. Dataset1: Insurance churn prediction-machine hackthan.

Classifier	LSTM+ AOA	RNN+ AOA	RBM+ AOA	RBFN +AOA	SBLSTM-RNN+AOA
Accuracy	90.11	88.23	89.11	91.27	**97.69**
True Churn	89.21	88.12	91.11	89.23	95.22
False Churn	11.40	11.33	12.21	11.33	8.34
Specificity	94.53	83.54	91.56	89.23	82.76
Precision	89.19	82.12	94.31	92.34	94.31

and various parameters were tested. Data pre-processed in this model performs better enough for the entire classification methods in comparison to a classification performed without the pre-processing. Among classifiers examined, LSTMs show the best Accuracy of 95.47 in dataset 1 and SBLSTM-RNN shows 96.22, and 96.81 in datset2 and 3 respectively greatest True churn of 94.11, SBLSTM-RNN achieving the most of highest accuracy values, which is underlined in bold letters. The accuracy comparison values are shown below in Figure 9 for better understandability.

E. PERFORMANCE OF DEEP LEARNING MODEL (SBLSTM-RNN) WITH DIMENSION REDUCTION(AOA)

Table10, Table 11, and 12 show the effectiveness of the Deep classification models with Dimensionality reduction (AOA) used, and various parameters were tested. This means that data pre-processed in this model performs better enough for the entire classification methods in comparison to a classification performed without the pre-processing. Among

TABLE 12. Dataset3: Health insurance lead prediction.

Classifier	LSTM+ AOA	RNN+ AOA	RBM+ AOA	RBFN+ AOA	SBLSTM-RNN+AOA
Accuracy	92.65	91.63	91.23	94.27	**97.89**
True Churn	90.23	91.11	92.98	90.32	94.11
False Churn	10.39	12.67	11.44	10.75	8.16
Specificity	94.35	82.45	91.65	88.22	81.67
Precision	89.19	88.21	91.31	89.43	95.21

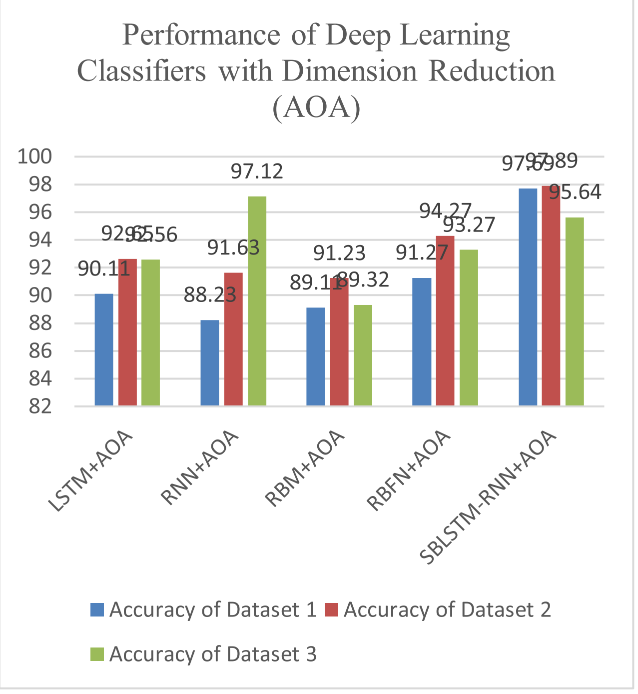

FIGURE 10. Performance of Deep Learning Classification algorithms with Dimension reduction (AOA).

classifiers examined, SBLSTM-RNN + AOA shows the best Accuracy of 97.87, and 97.69 in dataset2 and 1 respectively, and the greatest True churn of 95.22, which is underlined in bold letters. The accuracy comparison values are shown below in Figure 10 for better understandability.

In this section, the experimental validation of the SBLSTM-RNN-IDSA with AOA model is tested using a benchmark Insurance CCP dataset from the open source of Kaggle repository and it 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. 11 inspects the performance of the SBLSTM-RNN-IGSA with AOA model under distinct epoch count. The results demonstrated that the SBLSTM-RNN-IGSA with AOA model has effectively classified CR and NCR classes. For instance, with 200 epochs, the SBLSTM-RNN-IGSA with AOA model has identified 342 and 2844 samples under CR and NCR classes, respectively. In addition, with 400 epochs, the SBLSTM-RNN-IGSA with AOA methodology has identified 208 and 2845 samples under CR and NCR classes correspondingly. Moreover, with 800 epochs, the SBLSTM-RNN-IGSA with AOA technique has correspondingly identified 346 and 2843 samples under CR and NCR classes. Furthermore, with 1000 epochs, the SBLSTM-RNN-IGSA with AOA system has correspondingly identified 310 and 2844 samples under CR and NCR classes. With 1200 epochs, the SBLSTM-RNN-IGSA with AOA technique has correspondingly identified 371 and 2843 samples under CR and NCR classes.

FIGURE 11. Confusion matrix of SBLSTM-RNN-IGSA with AOA technique under distinct epoch count.

TABLE 13. Dataset2: Health-insurance-cross-sell-prediction.

Classifier	LSTM + AOA	RNN + AOA	RBM + AOA	RBFN+ AOA	SBLSTM-RNN+AOA
Accuracy	92.56	**97.12**	89.32	93.27	95.64
True Churn	91.32	90.66	91.89	91.53	93.91
False Churn	9.39	10.70	10.41	10.11	8.98
Specificity	93.11	88.12	92.19	88.87	81.10
Precision	89.10	84.14	94.24	95.73	93.29

Table 14 shows the effectiveness of the Deep classification models with Dimensionality reduction (SBLSTM-RNN-IGSA) used and various parameters were tested on multiple datasets, in each dataset our proposed model shows the best

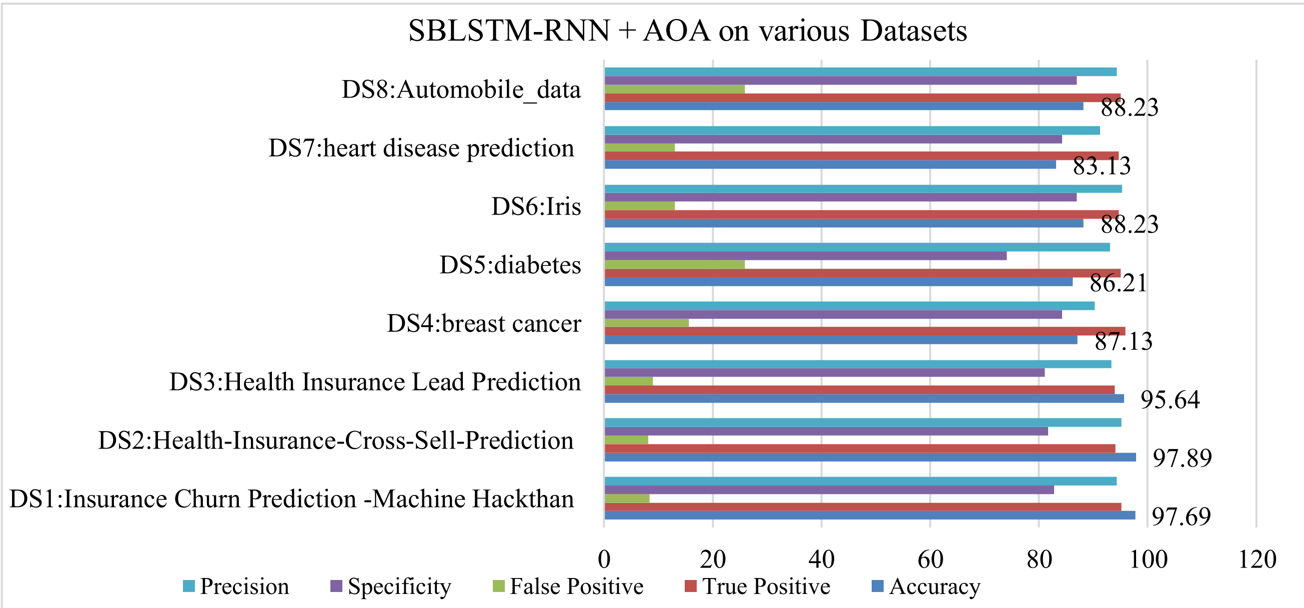

FIGURE 12. SBLSTM-RNN-IGSA + AOA on Various Datasets.

TABLE 14. Classifier: SBLSTM-RNN-IGSA with AOA.

	DS1: Insurance Churn Prediction -Machine Hack than	DS2: Health-Insurance-Cross-Sell-Prediction	DS3: Health Insurance Lead Prediction	DS4: breast cancer	DS5: diabetes	DS6: Iris	DS7: heart disease prediction	DS8: Automobile data
Accuracy	**97.69**	**97.89**	95.64	87.13	86.21	88.23	83.13	88.23
True Positive	95.22	94.11	93.91	95.89	95.02	94.69	94.69	95.02
False Positive	8.34	8.16	8.98	15.69	25.93	13.04	13.04	25.93
Specificity	82.76	81.67	81.11	84.31	74.07	86.96	84.31	86.96
Precision	94.31	95.21	93.29	90.23	93.03	95.35	91.23	94.34

accuracy values. The accuracy and other metrics comparisons are shown in the below Table14 and Figure 12 for better understandability. Our proposed model also suitable for other kinds of applications in to predict the values. The following are the datasets used in Table 14. DS1: Insurance Churn Prediction -Machine Hackthan, DS2: Health-Insurance-Cross-Sell-Prediction: Health Insurance Lead Prediction, DS4: breast cancer, DS5: diabetes, DS6: Iris, DS7: heart disease prediction and DS8: Automobile data. Table 15 consisting the accuracy information of various models on the Health Insurance cross cell prediction dataset, which was the dataset having highest accuracy.

Time spent executing instructions within a programme by the central processing unit (CPU) is referred to as CPU time. The amount of storage that an algorithm needs is referred to as its "space complexity." Table 15 presents a comparison of the various algorithms dimensionality reduction algorithms on three different datasets regarding their performance in terms of the amount of time taken by the CPU (measured in seconds), the amount of space required (measured in MB), and the number of iterations required for the algorithm to converge (speed of convergence). Due to the additional overhead of performing clustering before classification, the time required by the hybrid approach is greater than that required by the single approach. This is evident from the comparison. However, in order to achieve greater precision, this is a compromise that must be accepted. The proposed AOA algorithm gains good results as compared with other optimizers used in the comparison table 15.

F. PERFORMANCE OF DEEP LEARNING MODEL (SBLSTM-RNN-IGSA) ON HEALTH INSURANCE CROSS CELL PREDICTION DATASET

See Table 16.

TABLE 15. Time, Space and Speed of convergence for different optimization techniques.

Model		CPU Time (Seconds)	Memory (MB)	Speed of convergence(Sec)
1.Binary Particle swarm Optimization (BPSO)	Dataset-1	41.1	2828.9	14
	Dataset-2	54.2	2626.1	34
	Dataset-3	49.1	2726.4	24
2.Binary Gray Wolf Optimizer(BGWO)	Dataset-1	67.6	2594.3	66
	Dataset-2	77.6	2890.2	56
	Dataset-3	74.3	2578.3	67
3.Binary Crow Search Optimization(BCSO)	Dataset-1	58.2	2597.1	50
	Dataset-2	59.3	2793.4	59
	Dataset-3	56.2	2893.2	60
4.Binary Golden Eagle Optimization(BGEO)	Dataset-1	66.8	2659.9	53
	Dataset-2	62.2	2659.1	59
	Dataset-3	59.4	2769.3	43
5.Proposed Arithmetic optimization Algorithm(AOA)	Dataset-1	**33.24**	**2266.8**	**22**
	Dataset-2	**35.26**	**2443.3**	**32**
	Dataset-3	**37.14**	**2506.4**	**19**

V. CONCLUSION AND FUTURE WORK

In this work, Machine learning and Deep learning classifiers with and without feature extraction were applied. This paper to see if they could predict whether a customer would terminate their subscription and switch to another provider. This paper will be able to effectively anticipate customer turnover by comparing different classifiers while also addressing the key factor that contributes to client retention. The Arithmetic optimization algorithm(AOA) was initially conducted in pre-processing steps. The classifiers were then evaluated in a variety of ways. AOA will select some set of features to form the datasets and the selected optimal Features are supplied to classification models for finding churners and Non-Churners. Insurance customer churn prediction datasets are used in customer churn analysis and four sets of experiments are conducted. The results are examined in terms of various performance measures. To boost the classification efficiency of the SBLSTM-RNN model, improved gravitational search optimization algorithm (IGSA) based hyper parameter tuning process is involved in this study. it is discovered that SBLSTM-RNN+IGSA with AOA classifiers outperform others.Future work will necessitate the use of Deep Learning algorithms, which may improve forecasting rate accuracy. Finally, there is no time-series data in the datasets used in this study. Deep Learning techniques can be used in future research to implement time-series analysis. Reinforcement learning applied to time-series customer data is also promising. [44], [45], [46].

DECLARATIONS SECTION

Ethical Approval: Not Applicable

consent to participate: Not Applicable

consent to publish: Not Applicable

Acknowledgment: The author would like to thank Dr Nagendra Panini Challa, VIT-AP University, for assistance with technique and methodology for comments that greatly improved the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The Datasets described in this article are openly available in the open-source frame work and supporting files are available in

https://www.kaggle.com/mukulsingh/insurance-churn-prediction

https://www.kaggle.com/datasets/anmolkumar/health-insurance-cross-sell-prediction?+select=test.csv

TABLE 16. Comparison of Proposed SBLSTM-RNN-IGSA + AOA with other Existing Models.

Existing Techniques	Accuracy
Proposed SBLSTM-RNN-IGSA + AOA model	97.89
SVM-POLY using AdaBoost Praveen Asthana et al. [37]	97.00
Boosting proposed by Praveen Lalwani et al. [36]	81.71
Random forest proposed by Maria Spiteri et al. [35]	91.18
Deep & Shallow model by Rong Zhang et al. [34]	85.00
CRISP-DM logistic regression model by Andres et al. [33]	87.21
Hybrid GWO-KELM algorithm by Deepthi Das et al. [32]	95.00
RNN- ChOracle Model [38]	96.23
Ensemble learning-based models [39]	95.22
Random Forest (RF) algorithm [40]	88.63

https://www.kaggle.com/datasets/sureshmecad/health-insurance-lead-prediction

Conflicts of Interest: The authors declare no conflict of interest.

APPENDIX

Customer Churn Prediction (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-RNN model intends to proficiently forecast the occurrence of Customer Churn in the Insurance industry. Initially, the AOA model performs Pre-processing to transform the original data into a useful format. Besides, the SBLSTM-RNN model is employed to categorize data into churners and non-churners. To improve the CCP outcomes of the SBLSTM-RNN model, an optimal hyper parameter tuning process using Improved Gravitational Search Optimization Algorithm (IGSA) is developed.

It is difficult to study and understand the reasons for a customer to switch insurance providers when numerous forms of information are recorded from millions of clients. Insurance companies rely on the data to evaluate client behavior in order to minimize loss in an industry where retention of clients is equally essential with the earlier method being the costlier one. If insurance companies can anticipate whether or not a customer will switch, they can work to dissuade them from actually doing so. The ability to predict future attrition rates accurately is critical because it provides insight into potential profits for the company. Predicting the rate of customer attrition can also help your business identify and fix weak spots in customer support. Therefore, we will use the AOA-SBLSTM-RNN-IGSA Proposed Model to study churn detection in insurance policies using machine learning.

In this paper, the authors suggest using AOA-based feature extraction and a classification model-based Ensemble Deep learning to manage insurance client data. In this article, we used three insurance datasets to evaluate the efficacy of different configurations, as depicted in Figure 2. In this research, a novel AOA-SBLSTM-RNN model was created to improve the prediction of Customer Churn in the Insurance sector. In Frame work of CCP Model two phases are used.In Phase 1 the AOA model employs a procedure known as pre-processing to transform raw data into a more digestible format. Optimal subsets are selected from phase 1 from the datasets.In Phase 2, the SBLSTM-RNN model can distinguish between churners and non-churners in the data. The purpose of this work is to optimize the IGSA model's hyper parameters in order to produce better CCP results. The AOA-SBLSTM-RNN-IGSA method as a whole is depicted in Fig. 2.

This method for predicting customer churn employs a 4-stage setup.

1. The effectiveness of the base classification algorithms when dimension reduction is not applied.

2. The Effectiveness of Base Classification Algorithms Taking Dimension Reduction into Account.

3. The effectiveness of deep learning classification algorithms when dimension reduction is not applied.

4. The effectiveness of deep learning classification methods when dimension reduction is applied.

REFERENCES

[1] B. N. R. Chagas, J. A. N. Viana, O. Reinhold, F. Lobato, A. F. L. Jacob, and R. Alt, "Current applications of machine learning techniques in CRM: A literature review and practical implications," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Dec. 2018, pp. 452–458.

[2] A. S. Kumar and D. Chandrakala, "A survey on customer churn prediction using machine learning techniques," *Int. J. Comput. Appl.*, vol. 154, no. 10, pp. 1–4, 2016.

[3] Z. Li, "Science of customer management methods—How to reduce customer churn," *China Manage. Informationization*, vol. 1, no. 5, pp. 109–110, 2016.

[4] R. A. de Lima Lemos, T. C. Silva, and B. M. Tabak, "Propension to customer churn in a financial institution: A machine learning approach," *Neural Comput. Appl.*, vol. 34, no. 14, pp. 11751–11768, Jul. 2022.

[5] P. Granberg. (2020). *Churn Prediction Using Time Series Data*. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-294206

[6] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," *J. Big Data*, vol. 6, no. 1, pp. 1–24, Dec. 2019.

[7] A. De Caigny, K. Coussement, and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *Eur. J. Oper. Res.*, vol. 269, no. 2, pp. 760–772, Sep. 2018.

[8] F. Schena, "Predicting customer churn in the insurance industry: A data mining case study," in *Proc. Academy Marketing Sci. (AMS) World Marketing Congr. Rediscovering Essentiality Marketing*. Cham, Switzerland: Springer, 2015, pp. 747–751.

[9] L. Geiler, S. Affeldt, and M. Nadif, "A survey on machine learning methods for churn prediction," *Int. J. Data Sci. Anal.*, vol. 14, no. 3, pp. 217–242, Sep. 2022.

[10] V. V. Saradhi and G. K. Palshikar, "Employee churn prediction," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 1999–2006, Mar. 2011.

[11] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: A machine learning approach," *Computing*, vol. 104, pp. 271–294, Feb. 2022.

[12] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, and U. Abbasi, "Improved churn prediction in telecommunication industry using data mining techniques," *Appl. Soft Comput.*, vol. 24, pp. 994–1012, Nov. 2014.

[13] V. Umayaparvathi and K. Iyakutti, "A survey on customer churn prediction in telecom industry: Datasets, methods and metrics," *Int. Res. J. Eng. Technol.*, vol. 3, no. 4, pp. 1065–1070, 2016.

[14] E. Zdravevski, P. Lameski, C. Apanowicz, and D. Ślęzak, "From big data to business analytics: The case study of churn prediction," *Appl. Soft Comput.*, vol. 90, May 2020, Art. no. 106164.

[15] O. F. Seymen, O. Dogan, and A. Hiziroglu, "Customer churn prediction using deep learning," in *Proc. 12th Int. Conf. Soft Comput. Pattern Recognit. (SoCPaR)*. Cham, Switzerland: Springer, Apr. 2021, pp. 520–529.

[16] G. Pang, C. Shen, L. Cao, and A. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–38, 2020.

[17] N. Jajam and N. P. Challa, "Dynamic behavior-based churn forecasts in the insurance sector," *Comput., Mater. Continua*, vol. 75, no. 1, pp. 977–997, 2023.

[18] V. Umayaparvathi and K. Iyakutti, "Automated feature selection and churn prediction using deep learning models," *Int. Res. J. Eng. Technol.*, vol. 4, no. 3, pp. 1846–1854, 2017.

[19] M. J. Shabankareh, M. A. Shabankareh, A. Nazarian, A. Ranjbaran, and N. Seyyedamiri, "A stacking-based data mining solution to customer churn prediction," *J. Relationship Marketing*, vol. 21, no. 2, pp. 124–147, Apr. 2022.

[20] A. Amin, F. Al-Obeidat, B. Shah, M. A. Tae, C. Khan, H. U. R. Durrani, and S. Anwar, "Just-in-time customer churn prediction in the telecommunication sector," *J. Supercomput.*, vol. 76, no. 6, pp. 3924–3948, Jun. 2020.

[21] M. Scriney, D. Nie, and M. Roantree, "Predicting customer churn for insurance data," in *Big Data Analytics and Knowledge Discovery*. Bratislava, Slovakia: Springer, Sep. 2020, pp. 256–265.

[22] F. Shirazi and M. Mohammadi, "A big data analytics model for customer churn prediction in the retiree segment," *Int. J. Inf. Manage.*, vol. 48, pp. 238–253, Oct. 2019.

[23] L. Börthas and J. K. Sjölander, "Machine learning based prediction and classification for uplift modeling," Royal Inst. Technol. School Eng. Sci., Stockholm, Sweden, Tech. Rep. TRITA-SCI-GRU; 2020:002. MAT-E 2020:02, 2020. [Online]. Available: www.kth.se/sci

[24] 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," *Int. J. Forecasting*, vol. 36, no. 4, pp. 1563–1578, Oct. 2020.

[25] P. Kisioglu and Y. I. Topcu, "Applying Bayesian belief network approach to customer churn analysis: A case study on the telecom industry of Turkey," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7151–7157, Jun. 2011.

[26] C. Kirui, L. Hong, W. Cheruiyot, and H. Kirui, "Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining," *Int. J. Comput. Sci. Issues*, vol. 10, p. 165, Mar. 2013.

[27] P. K. Dalvi, S. K. Khandge, A. Deomore, A. Bankar, and V. A. Kanade, "Analysis of customer churn prediction in telecom industry using decision trees and logistic regression," in *Proc. Symp. Colossal Data Anal. Netw. (CDAN)*, Mar. 2016, pp. 1–4.

[28] J. Nagaraju and J. Vijaya, "Boost customer churn prediction in the insurance industry using meta-heuristic models," *Int. J. Inf. Technol.*, vol. 14, no. 5, pp. 2619–2631, Aug. 2022.

[29] A. B. W. Ali, "Prediction of employee turn over using random forest classifier with intensive optimized PCA algorithm," *Wireless Pers. Commun.*, vol. 119, no. 4, pp. 3365–3382, Aug. 2021.

[30] T. R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. K. R. Maddikunta, I.-H. Ra, and M. Alazab, "Early detection of diabetic retinopathy using PCA-firefly based deep learning model," *Electronics*, vol. 9, no. 2, p. 274, Feb. 2020.

[31] Z. Wu, L. Jing, B. Wu, and L. Jin, "A PCA-AdaBoost model for e-commerce customer churn prediction," *Ann. Oper. Res.*, vol. 2, pp. 1–18, 2022. [Online]. Available: https://doi.org/10.1007/s10479-022-04526-5.

[32] D. Das and R. R. Gondkar, "An effective machine learning model for customer attrition prediction in motor insurance using GWO-KELM algorithm," *Inf. Technol. Ind.*, vol. 9, no. 1, pp. 91–105, Feb. 2021.

[33] K. G. M. Karvana, S. Yazid, A. Syalim, and P. Mursanto, "Customer churn analysis and prediction using data mining models in banking industry," in *Proc. Int. Workshop Big Data Inf. Secur. (IWBIS)*, Oct. 2019, pp. 33–38.

[34] R. Zhang, W. Li, W. Tan, and T. Mo, "Deep and shallow model for insurance churn prediction service," in *Proc. IEEE Int. Conf. Services Comput. (SCC)*, Jun. 2017, pp. 346–353.

[35] M. Spiteri and G. Azzopardi, "Customer churn prediction for a motor insurance company," in *Proc. 13th Int. Conf. Digit. Inf. Manage. (ICDIM)*, Sep. 2018, pp. 173–178.

[36] N. Jajam and N. P. Challa, "Customer churn detection for insurance data using blended logistic regression decision tree algorithm (BLRDT)," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 1, pp. 72–83, 2023.

[37] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," *Simul. Model. Pract. Theory*, vol. 55, pp. 1–9, Jun. 2015.

[38] A. Khodadadi, S. A. Hosseini, E. Pajouheshgar, F. Mansouri, and H. R. Rabiee, "ChOracle: A unified statistical framework for churn prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 4, pp. 1656–1666, Apr. 2022.

[39] A. A. Khalil, Z. Liu, A. Salah, A. Fathalla, and A. Ali, "Predicting insolvency of insurance companies in Egyptian market using bagging and boosting ensemble techniques," *IEEE Access*, vol. 10, pp. 117304–117314, 2022.

[40] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector," *IEEE Access*, vol. 7, pp. 60134–60149, 2019.

[41] [Online]. Available: https://www.kaggle.com/datasets/mukulsingh/insurance-churn-prediction

[42] [Online]. Available: https://www.kaggle.com/datasets/anmolkumar/health-Insurance-cross-sell-prediction

[43] [Online]. Available: https://www.kaggle.com/datasets/sureshmecad/health-Insurance-lead-prediction

[44] A. A. Q. Ahmed and D. Maheswari, "An enhanced ensemble classifier for telecom churn prediction using cost based uplift modelling," *Int. J. Inf. Technol.*, vol. 11, no. 2, pp. 381–391, Jun. 2019, doi: 10.1007/s41870-018-0248-3.

[45] M. K. Gupta and P. Chandra, "A comprehensive survey of data mining," *Int. J. Inf. Technol.*, vol. 12, no. 4, pp. 1243–1257, Dec. 2020, doi: 10.1007/s41870-020-00427-7.

[46] M. Zohdi, M. Rafiee, V. Kayvanfar, and A. Salamiraad, "Demand forecasting based machine learning algorithms on customer information: An applied approach," *Int. J. Inf. Technol.*, vol. 14, no. 4, pp. 1937–1947, Jun. 2022, doi: 10.1007/s41870-022-00875-3.

[47] R. C. Staudemeyer and C. W. Omlin, "Evaluating performance of long short-term memory recurrent neural networks on intrusion detection data," in *Proc. South Afr. Inst. Comput. Sci. Inf. Technol. Conf.*, Oct. 2013, pp. 218–224.

NAGARAJU JAJAM received the M.Tech. degree in computer science and engineering from Jawaharlal Nehru Technological University, Kakinada, Andhra Pradesh, India, in 2013. He is currently pursuing the Ph.D. degree with the Vellore Institute of Technology-Andhra Pradesh University, Amaravati, India. His current research interests include soft computing, machine learning, and deep learning.

NAGENDRA PANINI CHALLA (Member, IEEE) has been an Assistant Professor Senior Grade-Two with the School of Computer Science and Engineering (SCOPE), VIT-AP University, since March 2022. During the Ph.D. degree, he worked on post-digitization activities on palm leaf manuscripts and has a good number of published research articles. His current research interests include machine learning and artificial intelligence algorithms across insurance and medical industry sectors. He has delivered many lectures on his research area with IMS BHU, Kendriya Vidyalaya, Tirupathi, and SCSVMV. He has been a reviewer for various eminent journals and conferences across India and abroad.

KAMEPALLI S. L. PRASANNA, photograph and biography not available at the time of publication.

C H VENKATA SASI DEEPTHI, photograph and biography not available at the time of publication.