International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

An Improvised Email Spam Detection using FSSDL-ESDC Model

Mr. N.A.S. Vinoth[1*], Dr. A. Rajesh[2]

Submitted: 07/11/2022 **Accepted:** 12/02/2023

Abstract: Email is a commonly available communication technology used to share information among people via the Internet. But the drastic upsurge in email misuses/abuses has led to a rising quantity of spam emails in recent times. Spam email classification by the use of data mining and machine learning (ML) models has gained significant attention among researchers owing to the positive effect on saving Internet users. Different ML and feature selection (FS) techniques can be employed to design effective email spam detection and classification approaches. In this aspect, this paper devises a novel feature subset selection with deep learning-based email spam detection and classification (FSSDL-ESDC) technique. The FSSDL-ESDC technique encompasses two major processes namely tokenization and stops word removal. In addition, a feature selection approach based on fruitfly optimization (FFO) is used to find an optimum subset of characteristics. Furthermore, for the categorization of email spam, the bidirectional long short-term memory (BiLSTM) approach is applied. In order to boost the email spam detection performance of the BiLSTM model, grasshopper optimization algorithm (GOA) is applied to finely tune the hyper parameters of the BiLSTM model. The improved performance of the FSSDL-ESDC approach is shown by a rigorous simulation study. The experimental results demonstrated that the FSSDL-ESDC approach outperformed the other state-of-the-art procedures.

Keywords: *Email spams, Spam filtering, Machine learning, Feature selection, Classification, Metaheuristics*

1. Introduction

Since the vast number of Internet users has expanded dramatically, E-mail has become a popular and efficient means of communication. As a result, e-mail management has become a significant and crucial concern for businesses and people, as it is easily abused. Commonly, Spam is determined as transmitting of unsolicited large amount of emails i.e., e-mail that wasn’t requested by various receivers [1]. A general characterization of spam is limited to unsolicited commercial e-mails, doesn’t consist of non-commercial solicitations like religious political/pitches, although it is unsolicited, as spam. An E-mail has been the most popular way of spamming on the Internet. In 2013, it has been stated that over 70% of e-mails transmitted around the world were categorized as spam [2]. Spam filters that use training approaches to understand the syntax and structure of an e-mail message are common. [3]. Besides, to resolve the spam problems, several methods depending on the Bayesian classifications are recommended by the scientists. International data group [4] anticipated that worldwide e-mail traffic increases to 60 billion messages every day. It comprises sending a huge number of recipients exactly similar unwanted mailings. Spam, unlike lawful business e-mails, is usually sent without the recipients' express consent and typically employs a variety of techniques to bypass e-mail filters. Everyday E-mail user receives hundreds of spam messages with novel content, from novel address i.e.,

manually created using robot software [5]. Filtering spam with conventional approaches as black-white list (domain, IP address, mail address) isn’t possible [6]. It is also possible to categorise spam communications in order to determine thematic dependency from geographical location. (For instance, what is the most common topic of spam mails sent from certain countries?). Method of text classifying and clustering approaches are effectively employed to spam problems in previous years [7]. In recent times, Adaption of machine learning (ML) approach for detecting the changing pattern of the problems has become highly promising. The ML approach is now utilized with novel techniques like Blockchain [8] etc. Considerable works on spam e-mail filtering have been made by methods like DT, NBC, and NN. In order to tackle the challenge of increasing amount of unsolicited emails, various approaches for e-mail filtering are being placed in several commercial products. ML based filter is capable of adapting itself to the behaviour and changing pattern of the spammers [9]. Since spammers often tend to present subtle modifications and develop novel methods of spreading spam e-mails in the widest range, this is a crucial advantage. The static filter isn’t capable of detecting these

[1*] *Research Scholar, Department of Computer Science and Engineering, Vels Institute of Science Technology and Advanced Studies, Chennai, Tamil Nadu, India*
[1*] *Email: vinoth.nas89@gmail.com*
[2] *Professor, Department of Computer Science and Engineering, Vels Institute of Science Technology and Advanced Studies, Chennai, Tamil Nadu, India*
[2] *Email: arajesh.se@velsuniv.ac.in*

subtle modifications, finally fails against a slightly changed spam pattern.

This paper devises a novel feature subset selection with deep learning-based email spam detection and classification (FSSDL-ESDC) technique. The FSSDL-ESDC technique encompasses two major processes namely tokenization and stops word removal. In addition, the fruit fly optimization (FFO) based feature selection (FS) technique is employed to identify an optimal subset of features. Furthermore, to categorise email spam, the bidirectional long short-term memory (BiLSTM) approach is used. The grasshopper optimization algorithm (GOA) is used to fine-tune the BiLSTM model's hyperparameters in order to improve the model's email spam detection results. A thorough simulation study is performed to demonstrate the FSSDL-ESDC technique's better results.

2. Literature Review

This section performs a detailed review of existing email spam filtering approaches. In Ahmed [10], Based on the Firefly algorithm The use of FS algorithms to reduce the dimensionality of features and enhance the accuracy of spam e-mail classification is discussed. For every firefly, the feature is expressed in binary form; in particular, the features are converted to binary using the sigmoid function. The provided BFA investigates the space of optimum feature subsets, and feature selection is based on an FF, i.e., accuracy with NBC. Shuaib et al. [11] used the WOA, a meta heuristic optimization technique, to pick relevant characteristics in an e-mail corpus using rotating forest algorithms to determine whether or not the e-mail is spam. The rotation forest techniques were estimated after and before FS using WOA utilising the whole data set.

Zamir et al. [12] introduced an FSEDM that depends on novel and existing features of email dataset, i.e., extracted afterward preprocessing. Then, various supervised learning technique is employed on the presented features in association with FS methods such as Relief-F, data gain and gain ratio for ranking the major features and categorize whether the email is spam or not. In order to decrease the spam detection error, FS based technique was given with SCA method. In Pashiri et al. [13], feature vector is upgraded with the SCA for selecting an optimum feature to train the ANN approach. Singh [14] Intelligent water drop algorithms based on evolution and swarms have been suggested for e-mail spam categorization. The presented method is utilized by the ML classification method is called NB classifiers. Intelligent water drop techniques are used to create subsets, and NB classifiers are used to the subset to determine whether or not the emails are spam.

Nayak et al. [15] proposed a method of data science for SMD utilizing ML approach. A hybrid bagging method is utilized for detecting spam e-mails that implement the 2, NB, and J48 (viz., DT) ML methods. This algorithm is employed as input using a dataset i.e., separated into distinct sets with the data science. E-mail classifications could be achieved based on the patterns of repeated keywords and many other parameters such as Cc/Bcc, domain, header, and so on. i.e., presents in their framework. Saleh et al. [16], presented the efficacy of utilizing an NSA for detecting anomalies employed to filter spam messages. NSA has higher efficiency and a lower false detection rate. The developed method works intelligently via 3 detection stages for determining an e-mails legitimacy on the basis of knowledge collected in the training stage. The system works by eliminating NSA related to the functionalities of T cells in biological system. Akinyelu and Adewumi [17] reported the usage of RF-ML approach in classifications of phishing attacks, through the primary goal of emerging an enhanced phishing e-mail classifiers using an improved predictive accuracy and less amounts feature.

GuangJun et al. [18] introduced the application of ML based spam detection technique for detecting accurately. ML classifiers such as LR, KNN, and DT are used in this approach to classify ham and spam messages in mobile device transmission. This method is put to the test using SMS spam collection datasets. Fayaz and his colleagues [19] presented an ensemble ML method which integrates prediction from MLP, KNN, and RF also predicts the result of the reviews as real/spam (nonspam), depending upon the majority vote of contributing methods. Later, 3 distinct election methods were used for diminishing the feature space and filter-out the ten major optimum features.

3. The Proposed Model

The whole working procedure of the proposed FSSDL-ESDC model is shown in Figure 1. The FSSDL-ESDC approach was used to create a novel email spam detection and classification model in this study. Pre-processing, TF-IDF based feature extraction, FFO based FS, BiLSTM based classification, and GOA based parameter adjustment are all part of the proposed FSSDL-ESDC approach. The next sections go through the specifics of how these procedures function. The attributes used are, Message_id, Date, From, To, Subject, etc….

3.1. Pre-processing

At this stage, pre-processing takes place in two ways namely tokenization and stop word removal. It can be utilized for removing the redundant words and characters within all emails and generates the bag of words. The component ‘Count Vectorizer’ from Scikit-learn allocates numbers to all words/tokens. The sample was challenged for excluding English stop words such as A, In, The, Are, As, Is, and so on, but this would not be very useful in determining whether the email was spam or not. This example is part of the vocabulary study programme.

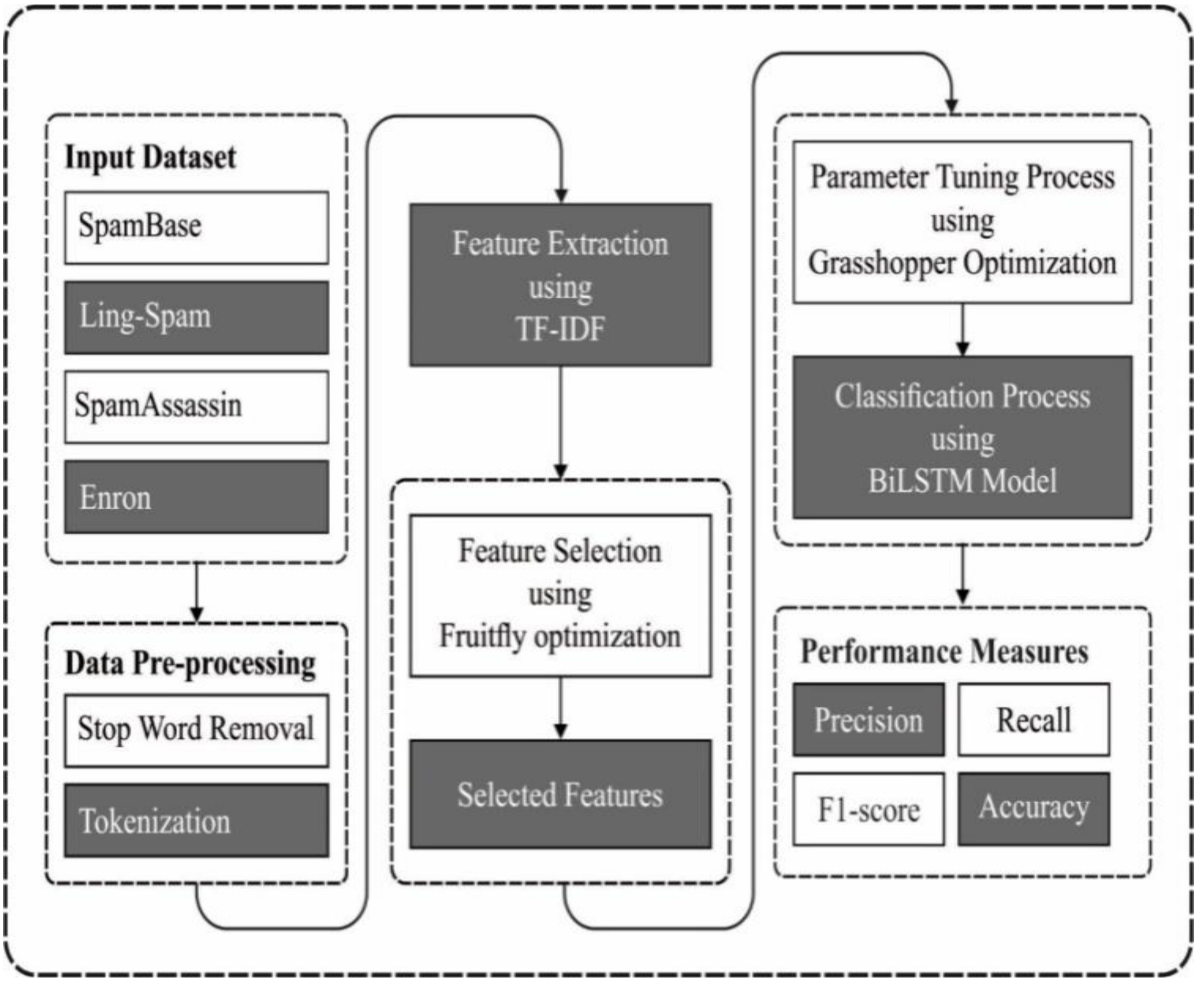
Input Dataset
SpamBase
Ling-Spam
SpamAssassin
Enron
Data Pre-processing
Stop Word Removal
Tokenization
Feature Extraction using TF-IDF
Feature Selection using Fruitfly optimization
Selected Features
Parameter Tuning Process using Grasshopper Optimization
Classification Process using BiLSTM Model
Performance Measures
Precision
Recall
F1-score
Accuracy

Fig. 1: Overall process of proposed model

3.2. Feature Extraction

Once tokenized, the proposed method implements 'Tfidf Transformer' component for computing the Inverse Document Frequency (IDF). In this study, a total of 57 features are extracted namely word_freq_make, word_freq_address, word_freq_all, word_freq_3d, char_freq_;,char_freq_(, char_freq_[, char_freq_!, etc. The most frequently occurring terms in texts are assigned values ranging from 0 to 1, and the lower the value of a word, the less likely it is to be unique. It allows methods or components for reading data to be used. [20]. The TF-IDF was computed by Eq. (1) where (t, doc) implies the term frequency (ter) within document (doc):

$$tf - idf(ter, doc) = tf(ter, doc) \times idf(ter) \quad (1)$$

where IDF was estimated by Eq. (2), provided n represents the number of documents:

$$idf(ter) = \log\left(\frac{\mathrm{n}}{\mathrm{df(ter)}}\right) + 1 \quad (2)$$

After this Process, the following features were given as an input to Feature Selection process namely, word_freq_make, word_freq_address, word_freq_all, word_freq_3d, char_freq_;,char_freq_(, char_freq_[, char_freq_!, etc.

3.3. Feature Selection

After feature extraction, the FS process is carried out using the FFO algorithm and the chosen 14 features are char_freq_!,word_freq_remove, word_freq_credit, char_freq_;, char_freq_(, char_freq_[, char_freq_$, word_freq_hp, capital_run_length_longest, word_freq_technology, word_freq_1999, word_freq_parts, word_freq_pm, and capital_run_length_total. FFO was identified as a novel approach derived from fruit fly foraging behaviour and used in interactive evolutionary modeling These kinds of creatures happen in maximum temperate and tropical climate zones globally. This species is effectual optical and olfactory sense if related to alternate creatures. It can be able to recognize varied aromas performed in air utilizing the olfactory organ while the feed source was located distant. Afterward, it flies near the food place utilizing the vision property. As a result, the following fruit fly food exploration model was created: (1) first, use the olfactory organ to smell the feed source; (2) Next, fly as close to the food supply as possible with keen vision; and last, determine the fruit fly's flocking place and proceed in the general direction. The flowchart of the FFO approach is shown in Figure 2. According to the food exploration rule, the FFO is divided into seven phases, as shown below.

Step 1: Set up the parameters: total progress, population size pop, and main fruit fly swarm location (X 0,Y 0).

Step-2: Initialized population [21]:

$$X_i = X_0 + rand,$$

$$Y_i = Y_0 + rand. \tag{3}$$

Step 3: Distance (D i) and odour (S i) evaluation:

$$D_i = \sqrt{X_i^2 + Y_i^2}, \tag{4}$$

$$S_i = \frac{1}{D_i}.$$

Step-4: Evaluation of fitness function (FF) (f_i) :

$$f_i = f(S_i). \tag{5}$$

Step-5: Recognize the minimal individual fruit fly with optimum FF (f_b) if related to other fruit flies:

$$[bestXbestindex] = \min\left(f(S_i)\right). \tag{6}$$

Step-6: Selection operation: Recollect the optimum FF and coordinate points (X_b, Y_b). Then, the fruit flies swarm move near the place with effectual FF in the application of sensitive visions:

$$f_b = bestX,$$

$$X_b = X\ (bestindex), \tag{7}$$

$$Y_b = Y\ (bestindex).$$

Step 7: Determine whether or not the end condition has been met. Else, go to Step 2; otherwise, stop the movement.

FS was regarded as binary optimization issue. It can utilize binary strings for representing the solution of FS issue. The vector consists of d items, where d is the number of features in the original dataset. If analogous functionality is available, put them to "1"; otherwise, set them to "0." The decision variable was defined as follows:

$$X = (x_1, x_2, \dots, x_d), x_i \in [0,1], i = 1,2,\dots,d. \tag{8}$$

FS is also regarded as multi-objective optimization issue. For maintaining the balance amongst the amount of FS and the classifier accuracy of solution, the fitness function (FF) was planned as:

$$Fitness = \alpha\gamma_R(D) + \beta\frac{|R|}{|C|} \tag{9}$$

$\gamma_R$implies the classifier error rate of provided classifier (BiLSTM method). αdenotes the importance of classifier accuracy, and βdenotes the importance of feature reduction. The total number of features is given by |C|, whereas the amount of FS is denoted by |R|.

After FF Process the chosen 14 features are, char_freq_!,word_freq_remove, char_freq_$, word_freq_credit, char_freq_;, char_freq_(, char_freq_[, word_freq_hp, word_freq_pm, capital_run_length_longest, word_freq_technology, word_freq_1999, word_freq_parts, and capital_run_length_total.

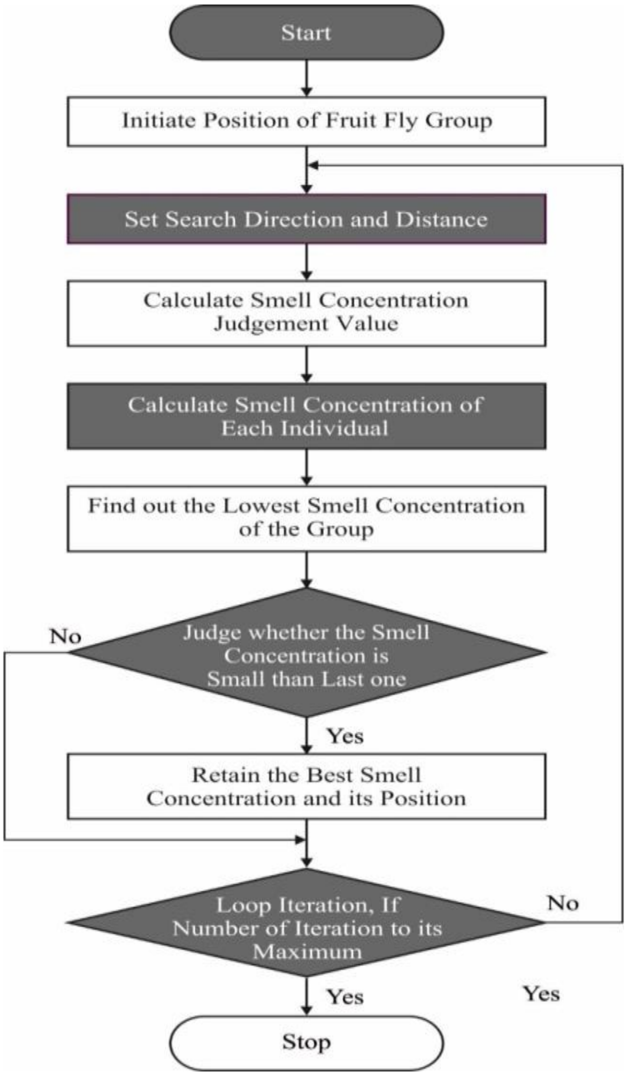
Start
Initiate Position of Fruit Fly Group
Set Search Direction and Distance
Calculate Smell Concentration Judgement Value
Calculate Smell Concentration of Each Individual
Find out the Lowest Smell Concentration of the Group
No
Judge whether the Smell Concentration is Small than Last one
Yes
Retain the Best Smell Concentration and its Position
Loop Iteration, If Number of Iteration to its Maximum
No
Yes
Yes
Stop

Fig. 2: Flowchart of FFO algorithm

3.4. Spam Classification

The chose feature subsets are passed into the BiLSTM model for spam classification purposes. Typically, an RNN was made up of problems learning long-term reliance. The LSTM-based approach was extended to RNNs that can handle decreasing gradient problems. An LSTM manner acquires vital features from input as well as maintain the data to huge time interval. The decision of removed or maintained the data is dependent upon weight value assigned to data in trained model. As a result, the LSTM approach learns the relevance of data for maintaining or eliminating it. Typically, an LSTM approach contains three gates: forget, input, and output. The framework of Bi-LSTM is shown in Figure 3.

Forget Gate. The sigmoid function was usually employed to make a decision which needs exits that deleting from LSTM memory. These decisions were basically dependent on $h_{z-1}$and x_t. As a result of this gate f_z, a value amongst [0,1], However, 0 denotes the scholarly value, whereas 1 denotes the total value. |C|. The resultant was estimated as:

$$f_z = \sigma(W_{f_h}[h_{z-1}], W_{f_x}[x_z], b_f) \quad (10)$$

where b_f , the constant which is identified as *bias* value.

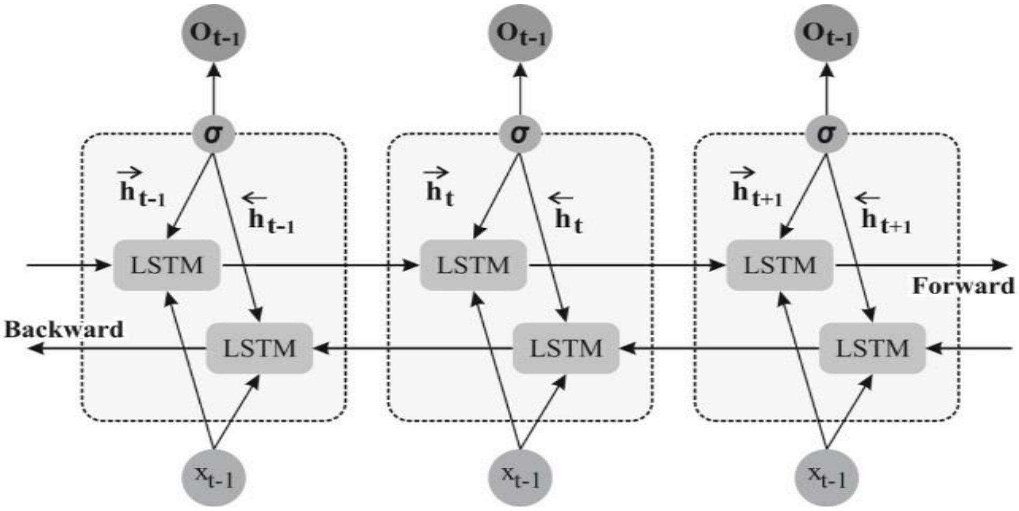

Figure 3: Structure of Bi-LSTM

Input Gate. It aims in determining whether or not new data is currently stored in the LSTM memory. These gates have two levels: sigmoid and tanh layers [22]. The sigmoid layer produces the values that need to be improved, while the tanh layer creates a vector of new candidate values that are less memory intensive for the LSTM. The results of two layers were calculated as follows:

$$i_z = \sigma\left(W_{i_h}[h_{z-1}], W_{i_x}[x_z], b_i\right) \quad (11)$$

$$c_z = \tanh\left(W_{c_h}[h_{z-1}], W_{c_x}[x_z], b_c\right) \quad (12)$$

wherei_z denotes whether or not a value is required to be upgraded, and the vector of new candidate values is c_z which is relevant to LSTM memory. The combination of these two layers improves LSTM memory by forgetting the current value using the forget gate layer, multiplying the previous value ($for\ intance, c_{z-1}$) and then adding a new candidate value $i_z * c_z$. The subsequent formula represents their mathematical expression:

$$c_z = f_z * c_{z-1} + i_z * c_z \quad (13)$$

where f_z stands for the resultant of forgetting gate which the value amongst [0,1] where 0 represents the last time a value was rid; meanwhile, 1 denotes last time a value was kept maintained.

Output Gate. This gate makes a decision based on a sigmoid layer, resulting in the separation of LSTM memory. Then, to map the values between [-1, 1], it produces a nonlinear tanh function. The product was eventually amplified by sigmoid layers. The equations to estimate the outcome are represented by the following formula:

$$o_z = \sigma\left(W_{o_h}[h_{z-1}], W_{o_x}[x_z], b_o\right) \quad (14)$$

$$h_z = o_z * \tanh(c_z) \quad (15)$$

whereo_z refers the resultant value, and h_z demonstrates the value amongst [-1, 1].

3.5. Hyper parameter Optimization

For optimally tuning the learning rate in the BiLSTM model, the GOA is employed in such a way that the classification performance gets improved. Grasshoppers were creepy and separated as bugs. Plant harvesting is usually a failure since it uses all plant crops. In the larval stage, the swarm nymph is heavily used. An essential place of grasshopper has demonstrated in Eq. (16).

$$G_i = Soc_i + Gravity_i + Wind_i \quad (16)$$

whereSoc_i refers the social communication, $Gravity_i$ gravitational pull on i^{th} grasshoppers, and $Wind_i$ denotes wind advection. For resolving the grasshopper functions, Some functions were evolved from social communications, gravitational forces, and wind advection to resolve the grasshopper functions.

Social interaction: It may not be possible to generate diverse solid energy among grasshoppers with more isolation due to limited advantages. These problems are solved utilizing the division amongst grasshopper and mapped in terms of 1 and 4, so, the social interaction was illustrated as:

$$Soc_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} S\left(p_{ij}\right)\hat{p}_{ij} \tag{17}$$

So, $\hat{p}_{ij} = \frac{q_j - q_i}{p_{ij}}; p_{ij} = |q_j - q_i|$ where p_{ij} indicates the distance from ith and jth grasshoppers, *Soc*signifies the function that determines the performance of social force and $\hat{p}_{ij}$ refers the unit vectors from ith to jth grasshoppers.The number of grasshoppers is denoted by the letter N. The social force is defined by the function S, which is resolved by the implemented function:

$$Soc_{force} = fe^{-\frac{k}{l}} - e^{-k} \tag{18}$$

where$f$ implies the intensity of attractions, l represents the attractive length scale, and their capability was showcased.

1. **Gravity force and Wind advection:** The grasshopper's gravitational force ($Gravity_i$) controlled by using Eqs: (19) and (20). The nymph grasshopper doesn't have wings, and the utilization exceeds the wind way [23].
2. $Gravity_i = -gr\_con_g$ (19)
3. $Wind_i =$ $zlgr\_drift$ (20)

where$g$ indicates the gravitational constant, gr_con signifies the unity vectors,The continuous drift is denoted by l, and the unit vector near the wind direction is denoted by gr - drift. For overcoming the optimized issues, the stochastic approach is applying exploration as well as exploitation to elected accurate estimate of global optimal that is exhibited as:

$$G_i = \sum \left\{ S\left(|q_j - q_i|\right) \frac{q_j - q_i}{p_{ij}} - gr\_con_g + zlgr\_drift \right\} \tag{21}$$

In GOA, it can be making sure that grasshopper with optimum objective esteem has regarded as fittest grasshopper. The numerical approach was effective with exclusive parameters for exploration as well as exploitation in diverse optimization stages.

4. Performance Validation

The proposed model is tested using four datasets and the results are simulated using the Python programme. Table 1 provides the dataset's specific information. All four datasets contain instances under two classes namely SPAM and HAM (non-spam).

Table 1: Dataset Description

Dataset	Total Instances	SPAM	HAM	% of Spam	Year
Spambase	4601	1813	2788	39.40%	1999
Ling-Spam	2893	481	2412	17%	2000
SpamAssasin	6047	1897	4150	31%	2002
Enron	36715	20170	16545	55%	2006

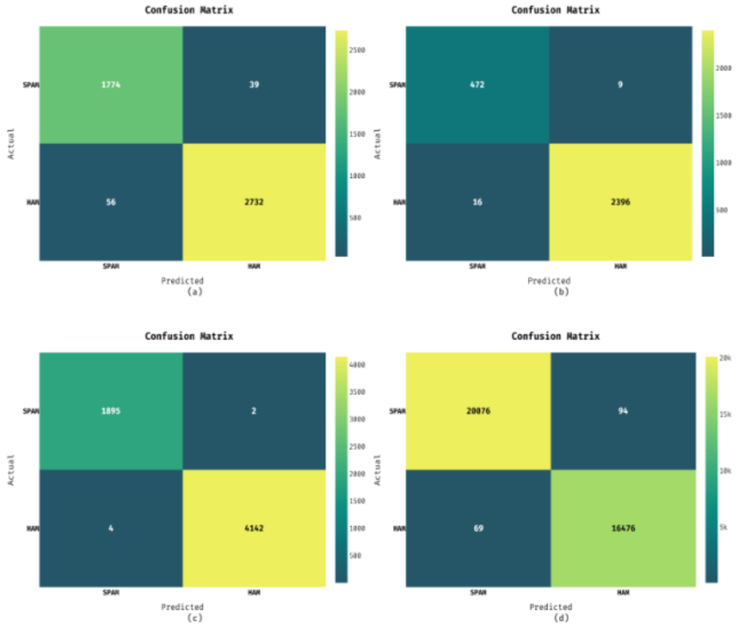

Fig. 4: Confusion Matrix a) Spambase Dataset b) Ling-Spam Dataset c) SpamAssasin Dataset d) Enron Dataset

On the four datasets used, Figure 4 displays the collection of confusion matrices generated by the FSSDL-ESDC approach. The FSSDL-ESDC approach categorised 1774 instances into the SPAM class and 2732 instances into the HAM class using the Spambase dataset. In addition, on the applied Ling-Spam dataset, the FSSDL-ESDC approach has classified 472 instances into SPAM class and 2396 instances into HAM class. Moreover, on the applied SpamAssasin dataset, the FSSDL-ESDC manner has classified 1895 instances into SPAM class and 4142 instances into HAM class. Furthermore, the FSSDL-ESDC method categorised 20076 cases into the SPAM class and 16476 instances into the HAM class on the Enron dataset.

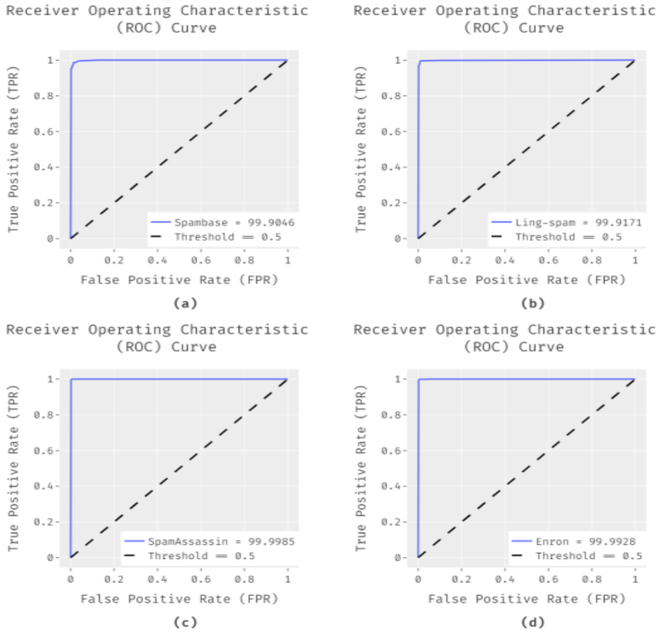

Fig. 5: ROC analysis of proposed model

Figure 5 shows the set of ROC curves accomplished by the proposed model on the applied four datasets. The results showcased that the proposed model has resulted in an increased ROC of 99.9046, 99.9171, 99.9985, and 99.928 Spam base, Ling-spam, Spam Assassin, and Enron datasets respectively.

Table 2 and Figure 6 shows the FSSDL-ESDC technique's email spam classification results on four datasets. The experimental findings demonstrated that the FSSDL-ESDC approach produced effective results on all datasets used.

Table 2: The Proposed FSSDL-ESDC Model's Results on the Applied Dataset

Datasets	Metrics			
	Precision	Recall	Accuracy	F-Score
Spambase	0.9694	0.9785	0.9794	0.9739
Ling-spam	0.9672	0.9813	0.9914	0.9742
SpamAssassin	0.9979	0.9989	0.9990	0.9984
Enron	0.9966	0.9953	0.9956	0.9960

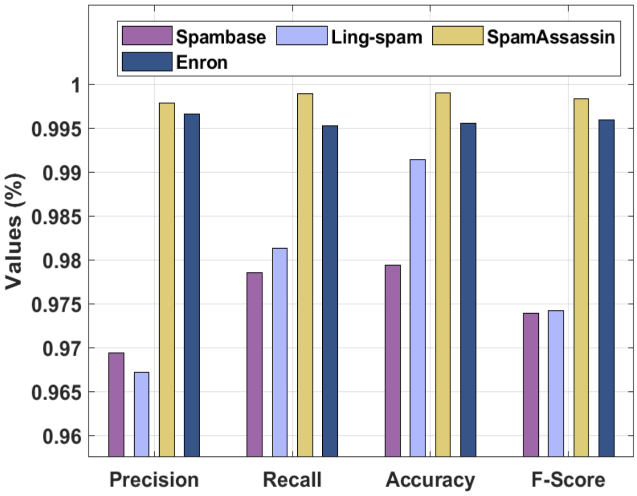

Fig. 6: FSSDL-ESDC Model Result Analysis with Different Measures

The Comparison research is done in Table 3 to demonstrate the improved performance of the FSSDL-ESDC approach. It demonstrated that the Base-RF and PSO-RF techniques have obtained an ineffective classification outcome over the other techniques. At the same time, the GSCV-DT, RSCV-DT, PSO-DT, Base-DT, and GA-DT techniques have gained a slightly improved outcome over the earlier methods.

Table 3: Analysis of Existing and Proposed Results: FSSDL-ESDC Model

Methods	Spambase	Ling-spam	Spam Assassin	Enron
Base-SGD	0.9560	0.9782	0.9928	0.9912
Base-RF	0.8176	0.8722	0.8925	0.8088
Base-DT	0.9242	0.9256	0.9847	0.9605
PSO-SGD	0.9673	0.9811	0.9967	0.9920
PSO-RF	0.8517	0.8945	0.9733	0.8845
PSO-DT	0.9213	0.9001	0.9933	0.9362
GA-SGD	0.9637	0.9828	0.9983	0.9921
GA-RF	0.9637	0.8726	0.9817	0.9405
GA-DT	0.9274	0.9190	0.9933	0.9607
GSCV-SGD	0.9794	0.9879	0.9950	0.9917
GSCV-RF	0.9564	0.9259	0.9833	0.9476
GSCV-DT	0.9176	0.9225	0.9883	0.9550
RSCV-SGD	0.9661	0.9879	0.9966	0.9914
RSCV-RF	0.9503	0.8984	0.9617	0.9427
RSCV-DT	0.9200	0.9363	0.9900	0.9493
Bi-LSTM	0.9749	0.9892	0.9986	0.9924
FSSDL-ESDC	0.9794	0.9914	0.9990	0.9956

Next to that, RSCV-RF, Base-SGD, GA-RF, RSCV-SGD, and PSO-SGD techniques have exhibited moderate performance. Furthermore, the Bi-LSTM and GSCV-SGD techniques have resulted in a competitive performance over the other compared methods. However, the proposed FSSDL-ESDC technique has accomplished superior performance with the maximum accuracy of 0.9794, 0.9914, 0.9990, and 0.9956 on the applied Spam base, Ling-spam, Spam Assassin, and Enron dataset respectively. From the above results analysis, it is ensured that the FSSDL-ESDC technique is found to be an effective tool for email spam classification.

5. Conclusion

In this study, a new email spam detection and classification model has been developed by the FSSDL-ESDC technique. Pre-processing, TF-IDF based feature extraction, FFO based FS, BiLSTM based classification, and GOA based parameter tuning are all part of the proposed FSSDL-ESDC approach. The FFO method is used to choose a subset of features, and GOA is used to modify the hyperparameters to achieve maximum classification performance. A thorough simulation study is performed to demonstrate the FSSDL-ESDC technique's better results. The simulation results showed that the FSSDL-ESDC approach outperformed the other state-of-the-art techniques. Clustering and outlier identification algorithms might be developed in the future to enhance email spam filtering results.

References

[1] S. Youn, 2014. SPONGY (SPamONtoloGY): Email classification using two-level dynamic ontology. *The Scientific World Journal*, *2014*.

[2] S. Youn and D. McLeod, A comparative study for email classification,*in Proceedings of the International Joint Conferences on Computer, Information, System Sciences, and Engineering (CISSE ’06), Bridgeport, Conn, USA, December* 2006.

[3] I. Androutsopoulos, G.Paliouras, and E. Michelakis, Learning to filter unsolicited commercial E-Mail NCSR “Demokritos,*Tech. Rep. 2004/2, March* 2004.

[4] S. Shankar and G. Karypis, Weight adjustment schemes for a centroid based classifier,*Computer Science Technical Report* TR00-035, 2000.

[5] R.MAlguliev,.,Aliguliyev, R.M. and Nazirova, S.A., 2011. Classification of textual e-mail spam using data mining techniques. *Applied Computational Intelligence and Soft Computing*, *2011*.

[6] M. L. Sang, S. K. Dong, and S. P. Jong, Spam detection using feature selection and parameters optimization,*in Proceedings of the 4th International Conference on Complex, Intelligent and Software Intensive Systems, (CISIS ’10)*, pp. 883–888, Krakow , Poland, February 2010.

[7] C. Paulo, L. Clotilde, S. Pedro et al., Symbiotic data mining for personalized spam filtering,*in Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technology, (IEEE/WIC/ACM)*, pp. 149–156, 2009.

[8] P. Cortez, C. Lopes, P. Sousa, M.Rocha, and M. Rio, 2009. Symbiotic data mining for personalized spam filtering. In *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology* (Vol. 1, pp. 149-156). IEEE.

[9] S.J. Murdoch, and R. Anderson, 2008. Tools and technology of Internet filtering. *Access denied: The practice and policy of global internet filtering*, *1*(1), p.58.

[10] B. Ahmed, 2020. Wrapper Feature Selection Approach Based on Binary Firefly Algorithm for Spam E-mail Filtering. *Journal of Soft Computing and Data Mining*, 1(2), pp.44-52.

[11] M.Shuaib, S.I.M. Abdulhamid, O.S. Adebayo, O. Osho, I. Idris, J.K.Alhassan, and N. Rana, 2019. Whale optimization algorithm-based email spam feature selection method using rotation forest algorithm for classification.

[12] A. Zamir,H.U. Khan, W. Mehmood, T.Iqbal, and A.U. Akram, 2020. A feature-centric spam email detection model using diverse supervised machine learning algorithms. *The Electronic Library*.

[13] R.T. Pashiri, Y.Rostami, and M. Mahrami, 2020. Spam detection through feature selection using artificial neural network and sine–cosine algorithm. *Mathematical Sciences*, 14(3), pp.193-199.

[14] M. Singh, 2019. Classification of spam email using intelligent water drops algorithm with naive bayes classifier. *In Progress in Advanced Computing and Intelligent Engineering* (pp. 133-138). Springer, Singapore.

[15] R. Nayak, S.A.Jiwani, and B. Rajitha, 2021. Spam email detection using machine learning algorithm. *Materials Today: Proceedings*.

[16] A.J. Saleh, A. Karim, B. Shanmugam, S. Azam, K. Kannoorpatti, M.Jonkman, and F.D. Boer, 2019. An intelligent spam detection model based on artificial immune system. *Information*, 10(6), p.209.

[17] A.A. Akinyelu, and A.O. Adewumi, 2014. Classification of phishing email using random forest machine learning technique. *Journal of Applied Mathematics*, 2014.

[18] L. GuangJun, S. Nazir, H.U. Khan, andA.UHaq, 2020. Spam detection approach for secure mobile message communication using machine learning algorithms. *Security and Communication Networks*, 2020.

[19] M. Fayaz, A. Khan,J.U. Rahman,A. Alharbi,Uddin, M.I. and Alouffi, B., 2020. Ensemble Machine Learning Model for Classification of Spam Product Reviews. *Complexity*, 2020.

[20] S. Gibson, B. Issac,L. Zhang, andS.M. Jacob, 2020. Detecting Spam Email with Machine Learning Optimized with Bio-Inspired Meta-Heuristic Algorithms. *IEEE Access*.

[21] Q.K. Pan, H.Y. Sang,J.H. Duan, andL. Gao, 2014. An improved fruit fly optimization algorithm for continuous function optimization problems. *Knowledge-Based Systems*, *62*, pp.69-83.

[22] F. Long, K. Zhou, andW. Ou, 2019. Sentiment analysis of text based on bidirectional LSTM with multi-head attention. *IEEE Access*, *7*, pp.141960-141969.

[23] H. Jia, Y. Li,C. Lang,X. Peng, K.Sun, and J. Li, 2019. Hybrid grasshopper optimization algorithm and differential evolution for global optimization. *Journal of Intelligent & Fuzzy Systems*, *37*(5), pp.6899-6910.