Egyptian Informatics Journal 26 (2024) 100473

Contents lists available at ScienceDirect

Egyptian Informatics Journal

journal homepage: www.sciencedirect.com

Full Length Article

Email spam detection by deep learning models using novel feature selection technique and BERT

Ghazala Nasreen [*], Muhammad Murad Khan, Muhammad Younus, Bushra Zafar, Muhammad Kashif Hanif

Department of Computer Science, Government College University, Faisalabad, Pakistan

ARTICLE INFO

Keywords:
Feature selection
Spam detection
BERT
LSTM
Machine learning

ABSTRACT

Due to the influx of advancements in technology and the increased simplicity of communication through emails, there has been a severe threat to the global economy and security due to upsurge in volume of unsolicited During the training of models, high-dimensional and redundant datasets may reduce the classification results of the model due to high memory costs and high computation. An important data processing technique is feature selection which helps in selecting relevant features and subsets of information from the dataset. Therefore, choosing efficient feature selection techniques is very important for the best performance of classification of a model. Moreover, most of the research has been performed using traditional machine learning techniques, which are not enough to deal with the huge amount of data and its variations. Also, spammers are becoming smarter with technological advancement. Therefore, there is a need for hybrid techniques consisting of deep learning and conventional algorithms to cope with these problems. We have proposed a novel scheme in this paper for email spam detection, which will result in an improved feature selection approach from the original dataset and increase the accuracy of the classifier as well. The literature has been studied to explore the efficient machine learning models that have been applied by different researchers for email spam detection and feature selection to acquire the best results. Our method, GWO-BERT, has given remarkable results with deep learning techniques such as CNN, biLSTM and LSTM. We have compared our models with RF and LSTM and used dataset: "Lingspam," which is a publicly available dataset. With different experiments, our technique, GWO-BERT, obtained 99.14% accuracy, which is almost equal to 100 percent.

1. Introduction

Emails have become a significant part of our lives in this era of advanced technology. Among common communication methods, email is the most economic and effective method now and helps people share a lot of significant information. Email was introduced in the mid-1990 s, and it changed our lives in every field, such as business, education, research, etc. Email has several advantages and also has the drawback that it fills our inbox daily with several known and unknown. All of us receive legitimate and spam emails, and we wish to avoid spam emails because they are problematic. According to the latest research, 56.87 % of the emails are spam emails from global traffic, which results in much of the bandwidth waste. It has been forecasted that in 2025, the daily email amount will reach up to 376.4 billion, which will also increase the number of spam Studies show that this increasing amount of spam emails will result in an increase in time, cost, and storage for business activities. As far as humans are concerned, this will disturb their mental health as well. Such problems make email spam detection a very considerable issue, and it should be resolved with effective solutions. Popular machine learning techniques are being used to solve many issues in every field of life, ranging from medical solutions to email spam detection. See (Fig. 1).

There is an increase in unauthorized access to devices through spam emails [1]. Organizations set up different filtering approaches by setting the rules and firewall configurations for unsolicited email detection. In this case, Google offers 99.9 % success in spam mail detection [2]. Spam emails are deployed in different areas, such as on cloud-hosted applications, on the gateway (router), or on the computer of a user. In order to resolve the spam detection issue, different machine learning techniques have been applied by different researchers, such as content-based

[*] Corresponding author.
E-mail addresses: miscathwal@gmail.com (G. Nasreen), muhammadmurad@gcuf.edu.pk (M. Murad Khan), myounas@gcuf.edu.pk (M. Younus), bushrazafar@gcuf.edu.pk (B. Zafar), mkashifhanif@gcuf.edu.pk (M. Kashif Hanif).

https://doi.org/10.1016/j.eij.2024.100473
Received 6 January 2024; Received in revised form 27 March 2024; Accepted 20 April 2024
Available online 30 April 2024

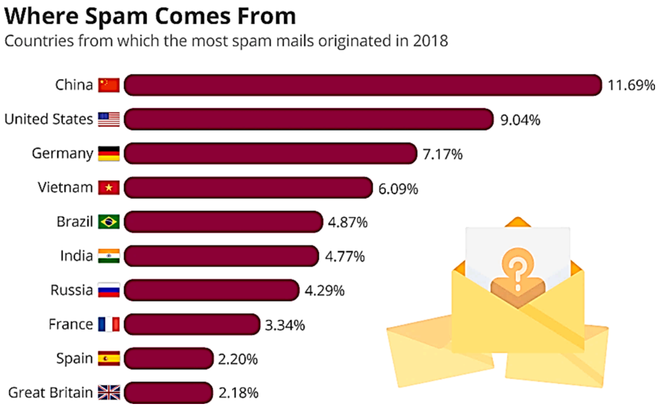

Fig. 1. Email spam ratio worldwide.

filtering or rule-based filtering. The rules for spam detection have been set, unlike 'knowledge engineering', and these are updated constantly manually, which is time- and resource-consuming as well. But the machine learning technique learns easily to recognize unsolicited emails and legitimate emails automatically.

The research was begun in 1970s in area of the feature selection that is significant for pattern recognition and data mining. Feature selection selects the optimal sub-set of the features among all the features in the original dataset by eliminating redundant and irrelevant features. This is necessary to enumerate as well as calculates all the possible subsets of features for optimal feature subset identification. All the possible sub sets from features are included in search space, where the size of search space will be 2n, and n here depicts total number of preliminary features in original dataset [3].

1.1. Filter- based feature selection approaches

In this type all the features are assigned a rank according to statistical and probabilistic properties without any classifier or learning method. These approaches are fast and can be divided into multivariate and univariate algorithms [4]. Relevance of all the features is assessed independently according to specified criteria of statistics. All the features are evaluated individually and the correlation between features is ignored which reduce performance of classification. There has been presented many univariate algorithms including Gini index (GI), Gain ratio (GR) Laplacian score (L-Score) [5] Term variance (TV) [6], Information gain (IG) [7], Laplacian score (L-Score) [8], Rrelief [9] and the Fisher score (F-Score) [10,11].

1.2. Wrapper-based feature selection approaches

Wrapper based feature selection approaches give more promising results for prediction than filter based approaches. Only specified approaches are employed during the search stage in wrapper based approaches, so in complex classifications problems they are more expensive, computationally. There are two main types of wrapper-based approaches of feature selection called random methods and greedy methods. Greedy method of wrapper-based feature selection approach uses "Hill Climbing" strategy, where the initially selected features are added or removed sequentially, to initial features subset. Sequential forward selection and the sequential backward selection are the classical approaches of greedy search approaches, where features are searched sequentially in sequential search approaches and also incline to be trapped in local optimum. The randomness is united into its search method in the random search method to search optimal feature sub-set from available feature range. Whale optimization algorithm (WOA) [12], Random mutation hill-climbing [13], the Gravitational search algorithm (GSA)[14,15], Biogeography-based optimization BBO) [16], Simulated annealing (SA) [17], the Krill herd algorithm (KH) [18], Ant Colony Optimization (ACO) [19], Differential evolution (DE) [20], and Artificial Bee Colony (ABC) [21] are the examples of random search algorithms [22].

1.3. Embedded-based feature selection approaches

In embedded approaches the classification model is firstly trained with original feature of dataset and the measurement of correlation of every feature is performed by employing the results. Measurement of feature selection is embedded during process of training of the classification model [23] and many embedded-based approaches have been proposed recently. ElAlami [24] proposed embedded-based approach and they used in Artificial Neural Network (ANN) and in Genetic Algorithm (GA) to select optimal subset of features.

Sugumaran, Muralidharan and Ramachandran [25] developed another embedded based approach in which they implemented proximal support vector machine and decision tree (DT) for diagnosis of roller bearing. These approaches are more complex and it's hard to improve the results of classification model. So they used wrapper based. Mohammed et al. utilized different models such as: KNN, SVM, Naïve Bayes and, Rule based algorithms for unsolicited emails detection and developed a spam vocabulary and legitimate E-mails used during training and testing of dataset [26]. They used Python programming language in their experiment on Email-1430 dataset and concluded that Naïve Bayes worked as the best classifier than SVM. Wijaya and Bisri [27] presented hybrid algorithm, including Logistic Regression and Decision Tree with False Negative threshold, and their experiment remained successful with performance of Decision tree. They also compared the results with the prior researches. They obtained 91.67 % accuracy with 'Spambase' dataset.

This research paper answers following research questions:

RQ1. How does the fusion of optimization algorithm increases classifier's accuracy of deep learning techniques for classification of emails? And it may help classifier to distinguish more accurately the spam and ham emails?

RQ2. Which of the feature selection algorithm helps to diminish the selected features of email dataset with deep learning classifier?

RQ3. Which of the word embedding scheme increases the deep learning classifier's accuracy for email spam detection?

RQ4. Does the word embedding scheme may help to reduce the model's execution time?

1.4. Contribution

Keeping in sight the above research questions, this paper makes the following contributions. The suggested GWO-BERT method with deep learning based email filtering method, makes following contributions to the field:

- *Improved Accuracy:* The deep learning models in this method has a capacity to recognize the complex patterns and GWO algorithm improves the identification of anomalies, which leads to an improved accuracy as compared to the conventional techniques.
- *Fast processing:* This method helps to speed up email filtering by deep learning model, and parallel computing capabilities by Grey Wolf Optimization. So this approach helps in feature reduction and in fast filtration of the emails.
- *Scalability:* The proposed model with efficient algorithm provides the scalability contribution to the spam email detection by optimization of the parameters used in model enhancing the understanding of the context, capturing the sequential patterns, and ensuring the adaptability to the evolving techniques of the spam detection, as we can see in the Table 7 we have implemented proposed technique on CNN, LSTM and biLSTM models. The method provides accommodations of

large volumes of data, making a real-time filtering of emails possible for large organizations.

- *Increased security:* The proposed method identifies and prevents phishing attacks more successfully and safeguards the sensitive data.

2. Related work

Most of the internet users consider an email as significant way of business and personal connection of people. Now days volume of email has been increased lot, so the email categorization can avoid spam emails to save the time and to resolve many problems. This is the reason categorization of email has become interesting topic for their research. Ghulam Mujtaba et al. [28] have presented an article in which they have broadly reviewed published articles on email classification during 2006 to 2016. In this article they presented analysis in different five steps: (1) datasets used by researchers for email classification purpose (2) e-mail classification methods (3) application areas (4) feature space used (5) performance measure methods used. Issues, research challenges and gaps were presented in this article.. They presented that many researchers have used email classification in different fifteen application areas such as spam categorization in emails, the phishing emails, multi folder email's categorization etc. The article concluded that most of the features have been used by researchers in classification are header part, content of body, JavaScript of email and URL etc. Email management has become a considerable issue now, so different approaches have been used for email classification such as Naïve Bayes algorithm, statistical Bayesian etc. But these algorithms are based on techniques of artificial intelligent and they have many problems such as less efficiency, not handling the sarcasm and low accuracy as presented in an article by Akash Kumar Singh et al. [29]. In this article they presented a technique to build model by Fuzzy Artificial neural network (ANN), NLP and different (ML) machine learning techniques for email categorization using the predefined protocols. They used email data set of Gmail in this system and applied Fuzzy ANN algorithm as there is less work is needed in email categorization by this technique. They converted the extracted features into a numerical score by this algorithm and then arranged the values according to the fuzzy ranges to apply categorization of mails. According to the article they have produced better results with high values [29].

Email is considered as the most secure and trusted communication source for the purpose of data transmission. Unwanted emails are growing day by day and the volume of email is also increasing which the main issue is due to increase in internet users. So different filters have been used by different researchers to get rid of these unwanted massages. There are different spam detection techniques such as Knowledge-based technique, Clustering techniques, Heuristic processes, Learning-based technique etc. Hanif Bhuiyan et al. [30] presented a study of these techniques implemented by different researchers for this problem of email classification such as SVM, Naïve Bayes, Bayes Additive Regression, K-Nearest Neighbor and KNN Tree. They have compared and evaluated these techniques and concluded that Naïve Bayes and SVM have been used in most of the email filtration process an these techniques have effective outcomes as well as some loop holes are there for the researcher when they increase the performance. Manikandan and Sivakumar [31] presented an review article on machine learning algorithms, for the text classification. They studied different classification techniques such as SVM, DT, NB, Rocchio's algorithm, K-Nearest Neighbor, Decision rules Classification, Genetic Algorithm (GA) and Fuzzy Correlation with advantages, disadvantages and applications. They concluded that Naïve Bayes, SVM and K-NN is the most appropriate algorithms, for the text classification and may give very efficient but improvement is need to improve their optimal results. Esha Bansal and Kumar Bhatia [32] presented an article of comparative analysis, of the existing email classification methods and compared their performance. They concluded that there is a need of reduced training time for feature selection and need of ensemble-based techniques to boost the accuracy improvement. According to them, feature selection algorithm should be combined with ensemble based techniques for better efficiency of classifier. Visalakshi et al. [33] proposed an approach named "ensemble classifier for the email spam classification in Hadoop environment". In this approach a gradient boost ensemble technique was implemented with combination of NB and DT algorithms which improved classification accuracy. The model showed 80 % accuracy with NB and almost 80 % precision. While Gradient boosting algorithm result in 93 % accuracy and almost 92 % precision this improved the performance of classifiers. Shradhanjali and Toran Verma [34] have proposed an approach in which they used SVM for email classification and obtained 98 % accuracy. They used the steps such as preprocessing, features extraction, training with SVM and testing. Priti Kulkarni and Haridas Acharya [35] have compared different existing email classification techniques which used email header field as a feature for classification. They concluded that KNN and DT algorithms give better results with header features and classification results were outclass with Bayes net. Mohammed Awad and Monir Foqaha [36] proposed a technique in which they combined ANN called particles swarm optimization (PSO) and radial basis function neural networks (RBFNN) techniques. They used also SVD and K-NN to improve the width and the weight of RBFNN and the result was better accuracy. See (Table 8).

Belkebir and Guessoum [37] worked on Bee Swarm Optimization (BSO), genetic algorithm (GA), and SVM on an Arabic text data set. They performed experiments for automatic text classification. The BSO algorithm was developed based on swarm behavior of the bees, in which search area is divided into sections to explore. Feng et al. [38] presented a hybrid approach between NB and SVM by proposing a method to generate the hyperplane on SVM by given dimensions to reduce data points during training for feature sections. Gibson et al. [39] have used a machine learning approach to detect email spam by tuning through bio-inspired approaches. They used two algorithms, GA and PSO, on datasets as classifiers and concluded that the NB multinomial performed better than the GA algorithm and obtained 95.06 % accuracy for the Lingspam dataset. Idris et al. [40] have selected a human immune model named the NSA model to implement, which is a modified machine learning technique. They implemented local selection DE on NSA-DE and a LOF as a fitness function, which maximizes the distance of emails produced as spam and ham, and presented an improvement in detection accuracy of 95 %. Idris and Selamat [41] have developed conventional techniques, utilized NSA-PSO algorithms, and compared their technique with other conventional techniques. Their experiment showed that NSA-PSO performed better for email spam detection than NSA. Karim et al. [42] have developed an anti-spam structure that is based on an unsupervised approach with a multi-algorithm clustering method and achieved 94.91 % accuracy. Murugavel and Santhi [43] have worked on the robustness of spam filter approaches to reduce spam emails and proposed a scheme called MSSA to handle spam enriching email user information. Ouyang et al. [44] proposed a model for spam filtering that includes SYN packet features (filter-based), DNS blacklists, filters based on message content and filters based on message traffic characteristics. This model uses the Decision Tree algorithm for classification, and just the first three layers have been focused on in this model. Shuaib et al. [45] presented a meta heuristic optimization algorithm named as WOA for selection of significant features of email and a technique called the Forest technique for spam classification of mail. They used the 'Corpus' dataset. Sugumaran, Muralidharan, and Ramachandran [46] have used the TFIDF method for support vector machine (SVM) algorithm implementation. Their analysis indicates that there is a least percent error, which may be considered most accurate approach. Wijaya and Bisri [47] presented hybrid based algorithm in which they used two different classifiers on their dataset. This method combines logistic regression with a decision tree with a false negative threshold. They used the 'Spambase' dataset in their experiment and compared their results with prior research. Their experiment increased the performance of a decision tree classifier with 91.67 % accuracy.

2.1. Problem statement

Table 1 shows different Email spam detection issues, which have been tried to be solved by developing various approaches but there have been faced various challenges by they systems such as analysis of imbalanced datasets, less performance of existing classifiers for real time situations, high cost function of the virtual annealing and the low convergence precision. Mostly the work has been done on filtering from header and not suitable for complex and large data. The challenges mentioned above are source of motivation to use deep learning models by new researchers which are more efficient and suitable for large amount of data. Hence, the proposed model GWO-BERT has been adopted by developing multi-objective feature selection.

2.2. Baseline model

The baseline model which is used in our study is RNN based LSTM model [50]. This model uses tokenization and this scheme has been used for the short text dataset Lingspam. The accuracy of the existing model is 96 % and the execution time is in 19 sec. which is relatively higher. We have proposed embedding scheme for long text email spam classification which will also increase the accuracy of the model. The proposed model used pre-processing steps of the data, BERT word embedding technique and LSTM classifier on Lingspam dataset.

Table 1
Comparative summary of state-of-the-art approaches for Email spam classification.

References	Methodology	Features	Challenges/ Limitations
Sayed et al. [39]	Beyesian classifier and NN	It improves recall It detection accuracy	This model did not attain better classification performance
Belkebir and Guessoum [37]	BSO + SVM, NN	Optimized features	They used Arabic text dataset
Idris et al. [40]	NSA-DE	DE improved accuracy	It is unable to apply parallel hybridization on two utilized approaches
Idris and Salamet [41]	NSA-PSO	It solves complex problems and improves performance	It is unable to apply parallel hybridization on these utilized approaches
Ouyang et al. [44]	DT	It has less computational time It finds error efficiently	It is not suitable for real-setting
Feng et al. [38]	NB + SVM	Improved accuracy	They used Chinese dataset
Wijaya and Bisri [47]	LR + DT	Improved accuracy It reduces False Positive rate	It has false negative threshold This model cannot be applied on large data
Shuaib et al. [45]	WOA	It improves performance	
Gibson et al. [46]	GA and PSO	It predicted the optimal results It improved accuracy	They used restricted numerical dataset
Karim et al. [42]	Ensemble algorithm	It resulted in better accuracy	It cannot work on subject field and body of email for clustering
Murugavel and Santhi [48]	MSSCA	It categorizes spam easily It helps to avoid hacking	It cannot stop relevant spam to fall into inbox
Muhammadzadeh and Gharehchopogh [43]	MAMH	It is used for high dimension problems	It has high computational time
Ramparasal et al. [49]	TDIDF	It is easy for calculation	Its slow for some approaches

3. Proposed email spam detection methodology

The goal of this research is to develop a 2 stage hybrid deep learning model which can accurately detect emails as spam or ham. Our model is combination of deep learning and traditional algorithms. Our priority of the research is the optimizing of hyper parameters to ensure that the deep learning classifier can operate in real-time with high validation accuracy on a new dataset. Fig. 2 presents a proposed framework that is consist of preprocessing, feature extraction and feature selection in the first stage, while in its second stage the embedding steps to prepare input dataset for training of deep learning model and the classification of dataset. Where during learning process, input features of dataset are introduced in model through an input layer, processed across hidden layers and output layer classifies finally email into spam and ham classes. The in-depth explanations of all the stages have been given in sections that follow.

3.1. Dataset

In order to implement our proposed method, we have used Ling spam [51] dataset. By preparing these emails of dataset adequately, we can extract the essential aspects which often present in spam and ham emails, such as HTML tags, JavaScripts and appealing URLs to target interest visitors. This approach ensures compelling and robust analysis to improve email security and the user experience.

3.2. Split sentences

Each email sentence from, dataset is labeled with a tag but here we only we need sentence part during pre-processing stage of the data, so next we split email dataset in two main categories called sentence and label.

3.3. Data preprocessing

Data preprocessing is very essential for email filtering process, it helps to transform the raw text into standardized form which can be then easily analyzed. The techniques used include parsing of email body, tokenization, stemming and lemmatization, which splits the text into

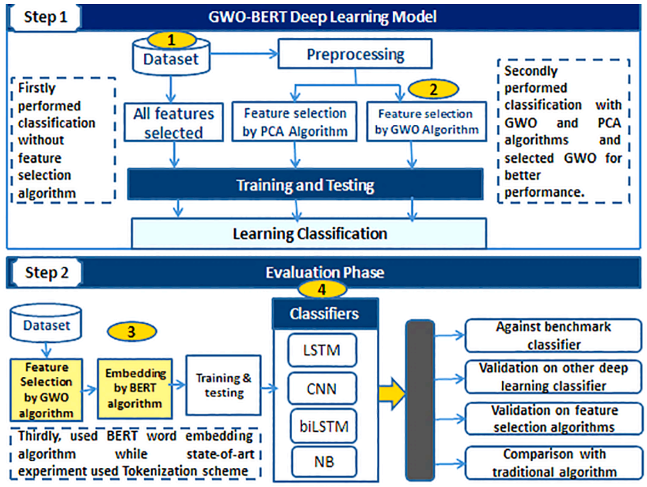
Step 1
GWO-BERT Deep Learning Model
1
Dataset
Preprocessing
2
Firstly performed classification without feature selection algorithm
All features selected
Feature selection by PCA Algorithm
Feature selection by GWO Algorithm
Secondly performed classification with GWO and PCA algorithms and selected GWO for better performance.
Training and Testing
Learning Classification
Step 2
Evaluation Phase
4
Classifiers
Dataset
3
Feature Selection by GWO algorithm
Embedding by BERT algorithm
Training & testing
LSTM
CNN
biLSTM
NB
Against benchmark classifier
Validation on other deep learning classifier
Validation on feature selection algorithms
Comparison with traditional algorithm
Thirdly, used BERT word embedding algorithm while state-of-art experiment used Tokenization scheme

Fig. 2. Proposed GWO-BERT Deep learning Methodology Architecture.

smaller parts, simplify the words and minimize the dimensionality and redundancy. So preprocessing helps to enhance overall effectiveness of the email filtering systems and make the text more manageable and consistent for its analysis. Data preprocessing stage has been divided into the following steps.

1. Stop words removal
2. Special symbol removal
3. Punctuation removal
4. Tokenization

3.4. Feature extraction

It's a crucial machine learning step which involves the identification of relevant raw data features, to improve performance of the model. In this experiment, we divided features into body based and the subject-line-based categories, which are numerical and Boolean features related to body of email, subject line and the specific words checklist. An effective feature extraction method can simplify modeling and improve the results.

3.5. Feature selection

Effective feature selection techniques help to simplify the models by removing irrelevant features from data which leads more complex models and longer processing times. Correlation feature selection (CFS), evaluates the subsets of features, on the basis of following hypothesis, 'Good features subsets contain the features highly correlated with the classification yet un-correlated to each other'. Following equation gives merit of feature subset ' S ' which consist of k features:

$$\text{Merit S} = \frac{krcf}{\sqrt{\text{k} + (\text{k} - 1)rff}} \tag{1}$$

We have used two standard methods called principal component analysis (PCA) and Grey Wolf Optimization (GWO). PCA method identifies essential features to reduce the dimensionality, and GWO finds the best combination of features for accurate predictions by model. Feature selection method reduces the training time and helps model to predict email classes faster. The details of both techniques are given below:

3.5.1. Principal component analysis (PCA)

This algorithm is a dimensionality reduction method that represents the high-dimensional data in lower dimensional space, preserving most of the information. This algorithm calculates 'k' principal components; by determining 'k' eigenvectors related to k which is largest eigenvalues of covariance matrix. This process includes steps such as standardizing data, computing covariance matrix, finding eigenvalues and eigenvectors, transforming data and returning a transformed data ' Y ' and eigenvectors. Some properties of the PCA include:

Property 1. For an integer q,$(1 \le q \le p)$ consider 'orthogonal linear transformation':

*y = B´**x**, where q is element vector and "B" is (q x p) matrix and let Σý= B´ ΣB is variance covariance matrix for "y" then Σ ý is denoted by tr (Σ y) maximized by taking (B = Aq) and where Aq is the first q column of A (B´is transpose of B).*

Property 2. Here consider again the orthonormal transformation, y = B′ **x** then tr **(**Σ y) can be minimized by taking (B =A^*q), where A_q is the last q column of A.

Property 3. *(Spectral decomposition of Σ)*

$$\Sigma = \lambda_1\alpha_1\alpha_1^{'} + \ldots + \lambda_p\alpha_p\alpha_p^{'} \tag{2}$$

3.5.2. Grey Wolf optimization (GWO)

It is *meta*-heuristics, swarm intelligence method. GWO is a meta-heuristic optimization algorithm inspired by means of the social hierarchy and hunting behavior of gray wolves. It is thought for its simplicity, efficiency, and ability to converge fast to highest quality answers. In the context of electronic mail unsolicited mail detection, GWO may be used to optimize the parameters of the BERT version, including mastering rates, weight decay, and other hyper parameters, to enhance its performance in classifying emails as unsolicited mail or not unsolicited ail. This algorithm uses few parameters and there is no source information is required during an initial search. GWO is a scalable, flexible, simple, and easy technique. During the search process, it has the special ability to strike a correct balance between exploitation and exploration, which results in favorable convergence. In GWO, when the search process starts, a grey wolf's random population is created. Then a group of four wolves is created, and positions are obtained to measure distances to the target prey. A candidate solution is represented by each wolf and updated by process of searching. The GWO algorithm performs very powerful operations that are controlled with only two parameters in order to maintain the exploration and exploitation and escape of stagnation of local optima. It has two functions: the rastrigin function and the sphere function. The function equation is:

Raster Function Equation:

$$\text{f}(x_1 \cdots x_n) = 10\,\text{n} + \sum\nolimits_{i=0}^{n}(x_i^{\,2} - 10\cos(2\pi x_i)) \tag{3}$$

Minimum at f (0, …,0) = 0 Rastrigin function is the challenging function for optimization problem with many cosine oscillations on a plane that bring together myriad of the local minima where particles may stuck.

3.5.3. Embedded-based feature selection approaches

In embedded approaches the classification model is firstly trained with original feature of dataset and the measurement of correlation of every feature is performed by employing the results. Measurement of feature selection is embedded during process of training of the classification model and many embedded-based approaches have been proposed recently. ElAlami [24] proposed embedded-based approach and they used in Artificial Neural Network (ANN) and in Genetic Algorithm (GA) to select optimal subset of features.

$$\text{f}(x_1 \cdots x_n) = \sum\nolimits_{i=0}^{n} x_i^{\,2} \text{minimum at f } (0 \cdots 0) = 0 \tag{4}$$

3.5.4. Bidirectional encoder representations from Transformers (BERT)

BERT embeddings are vectors which encapsulate word meanings, where the similar words have closer numbers in vectors. It has three separate input embeddings. To make a final token of input, the embeddings are brought together. The words are converted into embeddings to prepare data to work easily with model. It makes model enable to understand the semantic importance of word, in a numeric form and helps to perform all possible mathematical operations.

3.6. Classification

For classifying emails data we have used deep learning models (DL) such as LSTM, biLSTM, and CNN. We have also compared our results with Random Forest traditional machine learning model.

3.6.1. Long short term memory neural network (LSTM)

According to the literature review the traditional algorithms such as Logistic regression; NB etc. are unable to memorize past information or data. LSTM the variant of RNN is with ability to memorize past dataset and it can pass previously obtained information in a series form of networks just like architecture. Moreover, over time stamps it has strong gradient to hold long data sequences. It consists of three logic gates and memory gates. The Read gate is used to read data from memory cell, Write gate here writes data into the memory cell and Forget gate which deletes the old data. So the gates help LSTM networks to avoid from exploding and vanishing the gradients. LSTM has main advantage that it stores useful information and deletes unnecessary data. Model is mainly utilized for the data input in series form:

$$h_{t=}\ H(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \tag{5}$$

$$p_{t=}\ W_{hy}\,y_{t-1} + b_y \tag{6}$$

Eq. (1) and Eq. (2) represent computing equations, where x_t and y_t denote series inputs, ht represent the hidden memory cells, weight matrices by W and basics by b. Equations below represent hidden state of the memory.

$$i_{t=}\ \sigma(W_{ix} + W_{hh}h_{t-1} + W_{ic}C_{t-1} + b_i) \tag{7}$$

$$f_t\ = \sigma(W_{fx}x_t + W_{hh}h_{t-1} + W_{fc}C_{t-1} + b_f) \tag{8}$$

$$c_t\ = f_t\ {}^*C_{t-1} + i_t\ {}^*g(W_{cx}x_t + W_{hh}h_{t-1} + W_{cc}C_{t-1} + b_c) \tag{9}$$

$$o_t\ = \sigma(W_{ax}x_t + W_{hh}h_{t-1} + W_{oc}C_{t-1} + b_o) \tag{10}$$

$$h_t\ = \sigma_t{}^*h(c_t) \tag{11}$$

3.6.2. Bidirectional LSTM

This is the modification performed on normal LSTM networks [62] that enable it to give better performance for Natural Language Processing (NLP) issues.Here input is run in two ways, first from future to the past and second from the past to future. There is not a backward pass in Bidirectional LSTM and it obtains about 99.14 % accuracy.

Formula for calculation of the current state:

$$h_t\ = f(h_{t-1}\ , x_t \tag{12}$$

Where h_t is current state, h_{t-1} is previous state and x_t is input state.

Formula for applying activation function (tanh):

$$h_t\ = \tanh(W_{hh}h_{t-1} + W_{xh}x_t \tag{13}$$

3.6.3. CNN

CNN [63] model is consisted of finite number of processing layers which may learn different features of the input data such as images with multiple level abstraction. In initial the layers learn and then extract high-level features with lower level abstraction and then deeper layers learn, and extract a low level features, with a higher abstraction.

Convolutional formula:

$$(f^*g\ )(t)def = \sum\nolimits_{k=0}^{n} f(r)g(t-r)dr \tag{14}$$

A. Dropout: Main problem with a neural network is that when these are used to work with data of low then model over-fits data. This is why the data volume should be increased when used this network. We optimize the model by adding dropout to neural networks.

B. Dense: This is a fully connected layer which performs matrix multiplication operation on the input tensor and set of weights. It is followed by an activation function which is used to learn complex mappings between input and output data.

C. Rectified Linear Unit (ReLU): ReLU is a prevalent activation function of deep learning field. In the mathematical form it can be expressed as shown in Equation (1), where x is input to neuron.

$$f(x) = \max(0, x) \tag{15}$$

ReLU function has many advantages over other activation functions which include: computationally efficient and the fast convergence during training process. This function gives output zeros for all the negative inputs, results in fewer numbers of neurons being activated, and simplifies the model. Also this sparsity effect can help to prevent over-fitting problem of model.

D. Sigmoid: This is an activation function which takes the real numbers as input and binds output in [0, 1] range. Curve obtained of this sigmoid function is S shaped. Here is the mathematical representation of sigmoid given below:

$$\mathrm{s}(\mathrm{x}) = \frac{1}{1 + e^{-x}} \tag{16}$$

E. Adam Optimization: The Adam optimization is a widely used algorithm for training of deep learning models and has been built on the Stochastic Gradient Descent (SGD) method. This algorithm incorporates features of RMSProp and AdaGrad algorithms. Its advantage over SGD is that it can dynamically adjust learning rate, based on second moments of gradients. It helps in faster convergence and makes the model, computationally efficient. Its weight correction in gradient descent depends on slope shown in Eq. (3).

$$\text{Correction} = \alpha\frac{\delta \mathrm{i}}{\delta\theta i} \tag{17}$$

Eq. (4) evaluates the correction with momentum which considers weight parameter θ and cost function i. So this method uses momentum for enhancing optimization process.

$$\text{Correction} = \Upsilon^*\ \text{Previous Correction} + \alpha\frac{\delta \mathrm{i}}{\delta\theta i} \tag{18}$$

where ‘ γ’ is used to measure how much prior correction must have affected current corrections. Then the value of’ i ‘is updated and the suggested value of ‘ γ ‘ should be increased gradually from 0.5 to 0.9.

3.7. Model algorithm overview

The flow of this algorithm used in the above Fig. 2 has been presented here in the form of algorithm. There are following steps given:

1. An input layer convert words into the vectors
2. Split the data set
3. Data pre-processing
4. Feature selection by GWO algorithm
5. Apply word embedding technique BERT on processed data
6. Training and testing data
7. Classification by LSTM, biLSTM, CNN and NB

3.7.1. Proposed algorithm

The algorithm of proposed method for email spam classification is given here, in which we have used different steps from the beginning to the end to increase the performance of the classification through deep learning classifier. Here we have used feature selection algorithm GWO to select the best features out of data set and the embedding algorithm BERT will increase the accuracy of the model used.

```
BERT embedding
    Load the BERT model
    Tokenize preprocessed emails into BERT tokens
    Pad sequences to a fixed length
    Obtain BERT embeddings for each token
    Combine embeddings using techniques such as mean pooling or
    concatenation to get a fixed- size representation of each
Deep learning model
    Define the architecture of deep learning model
    Split the dataset into training, validation and test sets
    Train the model into training set: Input: BERT embeddings of
    Output: Binary classification (spam or not spam).
    Validate the model on validation set and fine tube hyper pa-
    rameters if necessary
```

```
        Evaluate the model on the test set to measure its performance such as accuracy, precision, recall, F1-score
        Optionally, perform model tuning using techniques such as dropout, batch normalization or adjusting learning rates
        Save the trained model for future use
    Post pre-processing
        Apply the trained model to classify new upcoming
        Optionally, further refine classification decision based on the confidence score or additional rules
        Take appropriate actions based on classification result such as move to spam folder, flag as potential spam
```

3.8. *Time complexity and space complexity of the proposed GWO-LSTM algorithm*

Space complexity and time complexity are very important measures for efficiency of the algorithm. Similarly, the time complexity is amount of time taken by an algorithm to execute as a function to length of an algorithm. Time complexity of a model depends on the hidden state size (H) and sequence length (T). Multiplication and element wise operations are performed which result in time complexity approximately GWO-LSTM is O (T*Hˆ2).

Space complexity is the measure of amount of memory space used by algorithm that is denoted by Big O notation. There is optimization process in GWO proposed algorithm which results potential difference in the training time. Addionally, there is lower space complexity of the GWO algorithm as compared to the existing traditional algorithm.

So performance of proposed algorithm is much better than existing algorithm in the aspect of accuracy, speed, space complexity and time complexity. Here from the results, we see that the optimization algorithm has decreased execution time, and increased the classification accuracy.

```
    Preprocessing:
        Tokenization: Split each email into words and sub words.
        Lowercasing: Convert all words into lower case to ensure consistency.
        Remove stop words: Remove common words that do not carry significant meanings.
        Lemmatization/Stemming: Reduce words to their base form to normalize variations.
        Feature extraction: Convert preprocessed text data into numerical representations.
    GWO (Grey Wolf Optimization) feature selection:
        Initialize the Grey wolf population
        Encode features into the chromosome of each wolf
        Evaluate the fitness of each wolf on the basis of selected features
        Update the position of wolves on the base of fitness
        Repeat until convergence or the maximum numbers of alterations reach
        Select the feature with the best fitness score
```

4. The experiments and results

We have carried out empirical experiments in order to analyze our proposed approach. Following subsections provide details about the data sets used in experiments and the methodology as well. The details of the experiments and the results have been provided also.

4.1. *Characteristics of datasets*

According to the literature review, there have been used many datasets for email spam classification task by different researchers. Recent articles have presented reviews of different kind of datasets which have been implemented in different researches, Table 4presents list of the datasets used.

As we can see from the Table 4 above, that datasets have been developed from 1998 to 2007 and no new dataset was developed after that. However, some articles show that most of the researchers have used Enron, Lingspam, Spamassassin, Spambase and PU datasets in their research as these are publically available. In this study we have used Lingspam and reasons for selecting this dataset are given below:

- Firstly, Lingspam dataset [51] has been utilized in many recent studies. Lingspam has been used in recent studies 2020 and 2022 so it means they are still in use in researches.
- Between 2000 and 2010 emails were generated with this dataset which characterizes change writing patterns and change in wordings in the emails from 10 year duration.
- Thirdly, this dataset has been used as it is domain specific and ham emails have been extracted from the scholarly linguistic discussions.

4.2. *Figure of merit*

All the experiments were conducted on system featuring 8th Gen Intel core i7, 3.6 GHz 8 12 GB RAM, Windows operating system. The software stack includes Python 3.7.0 and several packages such as matplotlib, numpy and pandas etc. Here the confusion matrix has been used to evaluate accuracy of classification model by presenting true positives, true negatives, false positives and false negatives. It helps in quick analysis of performance of model and identifies the areas for improvement.

$$True\ Positives(TP) : actual\ and\ predicted\ data\ point\ class\ is\ true. \tag{19}$$

$$True\ Negatives(TN) : actual\ and\ forecasted\ data\ point\ class\ is\ False. \tag{20}$$

$$False\ Positives(FP) : actual\ data\ point\ class\ is\ False\ and\ predicted\ class\ is\ true. \tag{21}$$

$$False\ Negatives(FN) : actual\ data\ point\ class\ is\ true\ and\ predicted\ class\ is\ False. \tag{22}$$

The formulations for the performance measures include:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + \text{FP} + \text{FN}} \tag{23}$$

$$\text{Precision} = \frac{TP}{TP + \text{FP}} \tag{24}$$

$$\text{Recall} = \frac{TP}{TP + \text{FN}} \tag{25}$$

$$\text{F} - \text{Score} = 2 * \frac{Precision*Recall}{Precision + Recall} \tag{26}$$

Table 2 above presents performance of different classification algorithms on specific data, assessed by different metrics such as: precision, recall, F-score, and accuracy. These algorithms tested in this experiment comprise of Deep Learning All Features, Deep Learning PCA Features, and Deep Learning GWO Features. See (Table 3).

Table 2
Summary of dataset used.

Dataset	Total Emails	Spams	Hams	Spam Rate	Year
Lingam	2894	481	2412	16.63	2002

Table 3
Computing Environment used for Experiments.

Python version	Operating system	CPU	RAM
3.7.6	Windows	8th Gen Intel core i7	12 GB

Table 4
Summary of different datasets used by other researchers for same problem.

Experiments	year	Algorithm	Dataset	Accuracy (%)
Drucker et al. [52]	1999	SVM	Emails	95
Banday and Jan [53]	2009	NB, KNN, SVM	Real life data set	95.6
DeBarr and Wechsler [54]	2009	RF	Custom collections	94.1
Shahi and Yadav [55]	2013	NB, SVM	Nepali SMS	91.64
Subramaniam et al. [56]	2010	NB	Gmail's emails	95.6
Eugene and Caswell [57]	2015	CNN	Enron	83
Abdulhamid et al. [58]	2018	ML	UCI	94.1
Du and Huang [59]	2018	LSTM	NLPCC2014 & Reuters	80.1
Lyubinets et al. [60]	2018	RNN	Arch Linux bug tracker & chromium bug tracker	87.1
Zhang, 2018 [61]	2018	CNN	Lingspam and Grumble	96
Salman et al. [50]	2022	LSTM	Lingspam	92

4.3. Empirical results

We have used 'Lingspam' email dataset in our experiments. Our baseline model is LSTM model which is without any feature selection algorithm. We have trained and tested our LSTM model with two feature selection algorithms PCA and GWO and observed the accuracies. GWO algorithm showed better results than PCA so we chose GWO for our next experiments. Next, we implemented proposed GWO-LSTM with different embeddings such as TD-IDF, w2v, BERT, Glove, where BERT performed best out of other embeddings and increased accuracy of the model also with fast training of dataset. We divided Ling spam email dataset in 80/20 ratio, with 80 % dataset for the training purpose and 20 % for testing. We performed total 13 experiments in three phases and the results have been shown in Table 5, 6 and 7 below:

4.3.1. Experiment 1

4.3.1.1. LSTM classifier with different feature selection algorithms implementation. In first experiment, we implanted LSTM without any feature selection algorithm and with all available features for classification with Lingspam dataset. The results have been presented in the Table 4 and Fig. 3 below demonstrates the performance of classifier across different key metrics. Accuracy of the model is impressive which 96 % with precision is and it indicates that model can classify most of test data correctly as shown in Table 5. In second experiment we implemented PCA feature selection algorithm with LSTM classifier on the same dataset and the accuracy was improved which is 98.53 %. Model achieved training and testing accuracy 98.89 % and 97.84 % respectively, so the feature reduction is clear here by PCA algorithm. In third experiment with GWO algorithm, feature selection the performance and accuracy have been increased of the model again. Our results show that model achieved 98.85 % accuracy. Training and testing accuracy of this model is 98.38 % and 97.19 % respectively, which are quite impressive. So here is the model's overall success but our focus is more on the accuracy improvement of the model so we have selected GWO-LSTM model for email spam classification. The graphs of the summary of accuracy and the loss of model have been presented in Figs. 3-5 below:

Table 5
Implementation of different PCA and GWO algorithms with existing LSTM and Lingspam dataset.

Model	Actual Accuracy	Training Accuracy	Testing Accuracy
Existing LSTM model	96 %	97 %	96 %
PCA-LSTM	98.53 %	98.89 %	97.84 %
Proposed GWO-LSTM	**98.85 %**	**98.38 %**	**97.19 %**

Table 6
Performance evaluation metric shows comparison of execution time of existing model and proposed GWO-BERT deep learning scheme.

Model	Accuracy of classification
K. Salman et al., 2022	96 %
Proposed scheme with GWO-BERT	98 %

Table 7
Implementation of proposed scheme GWO-BERT on different deep learning models.

Classifier	Accuracy	Precision	Recall	F-score
Existing CNN	92 %	95 %	92 %	91 %
Existing LSTM	96 %	94 %	96.5 %	94 %
TD-IDF-RF	93.89 %	94.69 %	96.85 %	95.72 %
GWO-BERT-CNN	97.28 %	94.16 %	97.28 %	96.11 %
GWO-BERT-LSTM	98.80 %	97.65 %	97.24 %	96.43 %
GWO-BERT-biLSTM	**99.14 %**	**99.89 %**	**94.73 %**	**97.29 %**

4.3.2. Experiment 2

4.3.2.1. Implementation of different embeddings on LSTM. In this phase we have implemented BERT embedding algorithms with proposed GWO algorithm with Lingspam dataset. The result shows that BERT has performed better than the tokenization used by the existing model. The accuracy is 98.89 % which is very good, which means that the model has classified emails more accurately. Similar the execution time of the model has been decreased also with BERT embedding which is 18.8 sec. So from the Table 6 we can see that the performance of BERT has been remarkable so we choose BERT for our proposed scheme GWO-BERT for deep learning models.

4.3.3. Experiment 3

4.3.3.1. Implementation of proposed scheme GWO-BERT on deep learning models. At the end for validation of our proposed scheme, we implemented proposed scheme on various deep learning models such as CNN, LSTM and biLSTM and the results have been displayed in Table 7 above. The results of existing CNN and LSTM have been given in Table 6 which is 92 % and 96 % accuracy respectively. On implementation of GWO-BERT on CNN we have achieved 97.28 % accuracy that is higher than existing CNN accuracy which was 92 %. Our proposed GWO-BERT scheme achieved 97.28 % accuracy which is a good improvement and it shows that it can classify emails more accurately than the existing LSTM model; similarly precision is 97.65 % that means that the proposed model can classify the emails into spam and ham with this precision. Similarly, the f-score of the model is also improved here which is 96.43 %. When we implemented GWO-BERT on biLSTM for comparison, there is more improvement in the results. We achieve 99.14 % accuracy

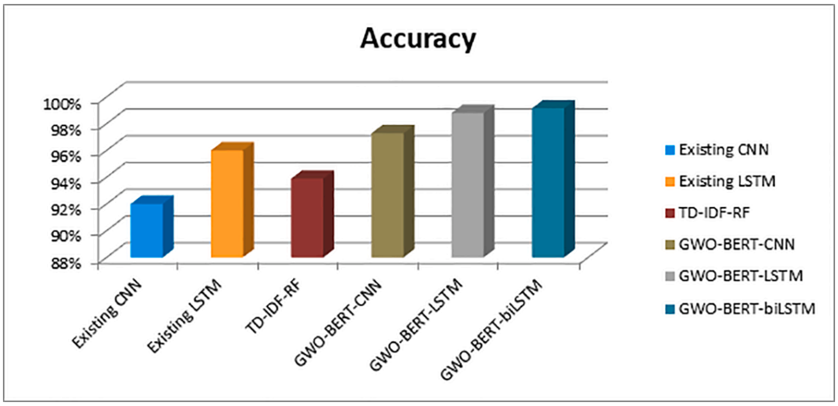

Fig. 3. Comparison evaluation of the results of different experiments with proposed GWO-BERT scheme.

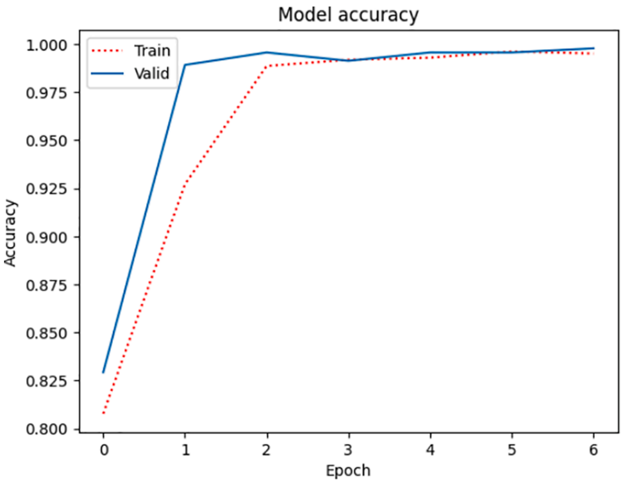

Fig. 4. Accuracy of GWO-BERT.

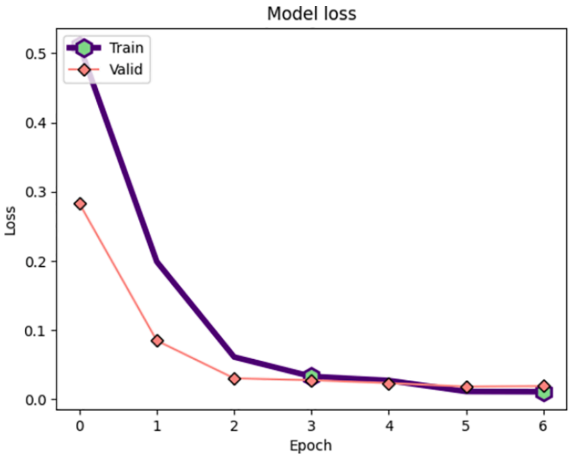

Fig. 5. Loss of GWO-BERT.

which is nearly equal to 100 % by BiLSTM model. Precision is 99.89 %, recall is 94.73 % and f-score is 97.29 %. The performance of GWO-BERT is remarkable with biLSTM and the accuracy is improved from 96 % to 99 %. We have also compared proposed scheme with traditional model Random Forest. There have been presented different graphs of the summary of accuracy and loss of different models below.

The Fig. 3 shows the comparison of GWO-BERT with other existing models; we can see the results and the improvement by our model. As our model has improved accuracy up to 99.14 % this is remarkable.

Here from the Table 9 the results can be checked where the accuracy of the model for spam detection has increased in a great sense. The accuracy of the existing model was 96 % and it has been improved to up to 99 % by using our proposed methodology.

5. Conclusion

An email is an effective, inexpensive and very fast method for the exchanging of information and messages through internet. Spam emails are very annoying to the end users and these are financially a damaging and may be source of security risk. Main objective of the spam emails is to gather the sensitive type of personal information of internet users. Moreover, the majority of the mails contain spams. In our study we have implemented various deep learning techniques such as CNN, biLSTM and LSTM models with GWO and BERT algorithms to classify Spam emails and legitimate Also we have compared proposed technique with the other different shallow techniques by implementing different (ML) machine learning algorithms. In this paper we have performed experiments which are based on combination of machine learning and (DL) deep learning algorithms and it indicates that using traditional feature selection algorithm with deep learning (DL) models has significantly increased accuracy of spam email detection rate which is from 96 % to 99.14 % and BERT word embedding helps in fast training of data and improves accuracy of the mode.

5.1. Future work and recommendations

We have implemented a few machine learning feature selection algorithms. In future work, other algorithms can be implemented with this model, and future work may be improved with the implementation of combination of (DL) deep learning classifiers which are based on image spam and texts. The biLSTM model can be used for more classes of email spam and similar problems. Our focus in future is to test our proposed model on a bigger size dataset. Also we are sure that by implementing hybrid and multi-model technique can increase the accuracy. In future we plan to work on the feature space.

Table 8
Implementation of different word embeddings on Lingspam dataset.

Word Embeddings	Training Accuracy (%)	Testing Accuracy (%)	Execution Time (sec)
GWO-TD-IDF	98.99	98.7	19
Proposed (GWO-BERT)	**99.91**	**99.14**	**18.8**

Table 9
Comparison of evaluation matrix of the proposed GWO-BERT with other state-of-the-art studies.

Author	Classifier	Dataset	Accuracy
AbdulNabi and Yaseen, 2021	CNN	Spambase, UCI	98.67 %
Bhopale and A. Tiwari, 2021	LSTM, RF	Enron	98 %
Tida and Hsu, 2022	CNN	Lingspam	97 %
Guo et al. 2022	SVM	Enron	96 %
Salman at al. 2022	CNN, LSTM	Lingspam	96 %
Proposed embedding GWO-BERT scheme	**LSTM**	**Lingspam**	**99.14 %**

CRediT authorship contribution statement

Ghazala Nasreen: Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Writing – original draft, Writing – review & editing. **Muhammad Murad Khan:** Investigation, Validation. **Muhammad Younus:** Investigation, Validation. **Bushra Zafar:** Formal analysis, Project administration. **Muhammad Kashif Hanif:** Formal analysis, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Luo H, Fang B, Yun X. A counting-based method for massive spam mail classification. Information Security Practice and Experience: Second International Conference, ISPEC 2006, Hangzhou, China, April 11-14, 2006. Proceedings 2 2006 (pp. 45-56). Springer Berlin Heidelberg.

[2] Dada EG, Bassi JS, Chiroma H, Adetunmbi AO, Ajibuwa OE. Machine learning for email spam filtering: review, approaches and open research problems. Heliyon 2019;5(6).

[3] Arif MH, Li J, Iqbal M, Liu K. Sentiment analysis and spam detection in short informal text using learning classifier systems. Soft Comput 2018 Nov;22:7281–91.

[4] Uesugi T. Toxic epidemics: agent orange sickness in Vietnam and the united states. Med Anthropol 2016 Nov 1;35(6):464–76.

[5] Huang J, Cai Y, Xu X. A hybrid genetic algorithm for feature selection wrapper based on mutual information. Pattern Recogn Lett 2007 Oct 1;28(13):1825–44.

[6] Koutroumbas K, Theodoridis S. Pattern recognition. Academic Press; 2008 Nov 26.

[7] Raileanu LE, Stoffel K. Theoretical comparison between the gini index and information gain criteria. Ann Math Artif Intell 2004 May;41:77–93.

[8] He X, Cai D, Niyogi P. Laplacian score for feature selection, in proceeding of Advances in Neural Information Processing Systems.

[9] Kira K, Rendell LA. A practical approach to feature selection. InMachine learning proceedings 1992 1992 Jan 1 (pp. 249-256). Morgan Kaufmann.

[10] Gu Q, Li Z, Han J. Generalized fisher score for feature selection. arXiv preprint arXiv:1202.3725. 2012 Feb 14.

[11] Tamoor M, Younas I. Automatic segmentation of medical images using a novel Harris Hawk optimization method and an active contour model. J Xray Sci Technol 2021 Jan 1;29(4):721–39.

[12] Mirjalili SM, Mirjalili SM, Lewis A. Grey Wolf Optimizer Adv Eng Softw 69: 46–61.

[13] Farmer ME, Bapna S, Jain AK. Large scale feature selection using modified random mutation hill climbing. InProceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. 2004 Aug 26 (Vol. 2, pp. 287-290). IEEE.

[14] Rashedi E, Nezamabadi-Pour H, Saryazdi S. GSA: a gravitational search algorithm. Inf Sci 2009 Jun 13;179(13):2232–48.

[15] Cheema A, Tariq M, Hafiz A, Khan MM, Ahmad F, Anwar M. Prevention techniques against distributed denial of service attacks in heterogeneous networks: a systematic review. Security and Communication Networks 2022 May;20(2022): 1–5.

[16] Du D. Biogeography-based optimization: Synergies with evolutionary strategies, immigration refusal, and Kalman filters.

[17] Wali A, Naseer A, Tamoor M, Gilani SA. Recent progress in digital image restoration techniques: a review. Digital Signal Process 2023 Aug;7:104187.

[18] Gandomi AH, Alavi AH. An introduction of krill herd algorithm for engineering optimization. J Civ Eng Manag 2016 Apr 2;22(3):302–10.

[19] Dorigo M, Birattari M, Stutzle T. Ant colony optimization. IEEE Comput Intell Mag 2006 Nov;1(4):28–39.

[20] Wu G, Mallipeddi R, Suganthan PN, Wang R, Chen H. Differential evolution with multi-population based ensemble of mutation strategies. Inf Sci 2016 Feb;1(329): 329–45.

[21] Karaboga D, Basturk B. Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems. InInternational fuzzy systems association world congress 2007 Jun 18 (pp. 789-798). Berlin, Heidelberg: Springer Berlin Heidelberg.

[22] Sheikh YA, Maqbool MU, Butt AD, Bhatti AR, Awan AB, Paracha KN, et al. Impact of rooftop photovoltaic on energy demand of a building in a hot semi-arid climate. J Renew Sustain Energ 2021;13(6).

[23] Nasreen G, Haneef K, Tamoor M, Irshad A. a comparative study of state-of-the-art skin image segmentation techniques with CNN. Multimed Tools Appl 2023 Mar;82 (7):10921–42.

[24] El-Alami FZ, El Alaoui SO, Nahnahi NE. A multilingual offensive language detection method based on transfer learning from transformer fine-tuning model. Journal of King Saud University-Computer and Information Sciences 2022 Sep 1;34 (8):6048–56.

[25] Tamoor M, Naseer A, Khan A, Zafar K. Skin lesion segmentation using an ensemble of different image processing methods. Diagnostics 2023 Aug 15;13(16):2684.

[26] Hameed B, Khan MM, Noman A, Ahmad MJ, Talib MR, Ashfaq F, et al. A review of Blockchain based educational projects. Int J Adv Comput Sci Appl 2019;10(10).

[27] Wijaya A, Bisri A. Hybrid decision tree and logistic regression classifier for email spam detection. In2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE) 2016 Oct 5 (pp. 1 4). IEEE.

[28] Mujtaba G, Shuib L, Raj RG, Gunalan R. Detection of suspicious terrorist emails using text classification: a review. Malays J Comput Sci 2018 Oct 24;31(4):271–99.

[29] Singh R, Bansal M, Gupta S, Singh A, Bhardwaj G, Dhariwal AD. Detection of social network spam based on improved machine learning. In2022 5th International Conference on Contemporary Computing and Informatics (IC3I) 2022 Dec 14 (pp. 2257-2261). IEEE.

[30] Sattu N. A study of machine learning algorithms on email spam classification (Doctoral dissertation, Southeast Missouri State University).

[31] Reddy KS, Reddy ES. An Efficient Methodology to detect spam in social networking sites. International Journal of Computer Science and Information Security (IJCSIS). 2017 Jul;15(7).

[32] Ali N, Fatima A, Shahzadi H, Ullah A, Polat K. Feature extraction aligned email classification based on imperative sentence selection through deep learning. Journal of Artificial Intelligence and Systems 2021 Aug 16;3(1):93–114.

[33] Renuka DK, Visalakshi P. Weighted-based multiple classifier and F-GSO algorithm for email spam classification. International Journal of Business Intelligence and Data Mining 2017;12(3):274–98.

[34] Verma T, Bhide S, Joshi S, Sharma A. EMAIL SPAM DETECTION.

[35] Kulkarni P, Saini JR, Acharya H. Effect of header-based features on accuracy of classifiers for spam email classification. Int J Adv Comput Sci Appl 2020;11(3).

[36] Foqaha MA. Email spam classification using hybrid approach of RBF neural network and particle swarm optimization. International Journal of Network Security & Its Applications 2016;8(4):17–28.

[37] Belkebir R, Guessoum A. A hybrid BSO-Chi2-SVM approach to Arabic text categorization. In2013 ACS International Conference on Computer Systems and Applications (AICCSA) 2013 May 27 (pp. 1-7). IEEE.

[38] Feng W, Sun J, Zhang L, Cao C, Yang Q. A support vector machine based naive Bayes algorithm for spam filtering. In2016 IEEE 35th International Performance Computing and Communications Conference (IPCCC) 2016 Dec 9 (pp. 1-8). IEEE.

[39] Gibson S, Issac B, Zhang L, Jacob SM. Detecting spam email with machine learning optimized with bio inspired metaheuristic algorithms. IEEE Access 2020 Oct;13(8): 187914–32.

[40] Ismaila I. Model and algorithm in artificial immune system for spam detection.

[41] Idris I, Selamat A. Improved email spam detection model with negative selection algorithm and particle swarm optimization. Appl Soft Comput 2014 Sep;1(22): 11–27.

[42] Karim A, Azam S, Shanmugam B, Kannoorpatti K. Efficient clustering of emails into spam and ham: the foundational study of a comprehensive unsupervised framework. IEEE Access 2020 Aug;17(8):154759–88.

[43] Mohammadzadeh H, Gharehchopogh FS. A novel hybrid whale optimization algorithm with flower pollination algorithm for feature selection: case study Email spam detection. Comput Intell 2021 Feb;37(1):176–209.

[44] Ouyang T, Ray S, Allman M, Rabinovich M. A large-scale empirical analysis of email spam detection through network characteristics in a stand-alone enterprise. Comput Netw 2014 Feb;11(59):101–21.

[45] Shuaib M, Abdulhamid SI, Adebayo OS, Osho O, Idris I, Alhassan JK, et al. Whale optimization algorithm-based email spam feature selection method using rotation forest algorithm for classification. SN Applied Sciences 2019 May;1:1–7.

[46] Sugumaran V, Muralidharan V, Ramachandran KI. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. Mech Syst Sig Process 2007 Feb 1;21(2):930–42.

[47] Sabah NU, Khan MM, Talib R, Anwar M, Arshad Malik MS, Ellyza Nohuddin PN. Google scholar university ranking algorithm to evaluate the quality of institutional research. Computers, Materials & Continua. 2023 Jun 1; 75 (3).

[48] Murugavel U, Santhi R. Detection of spam and threads identification in E-mail spam corpus using content based text analytics method. Mater Today: Proc 2020 Jan;1(33):3319–23.

[49] Aliero MS, Ghani I, Zainudden S, Khan MM, Bello M. Review on SQL injection protection methods and tools. Jurnal Teknologi 2015 Nov 12;77(13):49–66.

[50] Khan SA, Iqbal K, Mohammad N, Akbar R, Ali SS, Siddiqui AA. A novel fuzzy-logic-based multi-criteria metric for performance evaluation of spam email detection algorithms. Appl Sci 2022 Jul 12;12(14):7043.

[51] https://www.kaggle.com/datasets/mandygu/lingspam-dataset.

[52] Drucker H, Wu D, Vapnik VN. Support vector machines for spam categorization. IEEE Trans Neural Netw 1999 Sep;10(5):1048–54.

[53] Banday MT, Jan TR. Effectiveness and limitations of statistical spam filters. arXiv preprint arXiv:0910.2540. 2009 Oct 14.

[54] DeBarr D, Wechsler H. Spam detection using clustering, random forests, and active learning. InSixth conference on email and anti-spam. Mountain View, California 2009 Jul 16 (pp. 1-6).

[55] Shahi TB, Yadav A. Mobile SMS spam filtering for Nepali text using naïve bayesian and support vector machine. International Journal of Intelligence Science 2014 Jan;4(01):24–8.

[56] Khan MM, Bakhtiari M, Bakhtiari S. An HTTPS approach to resist man in the middle attack in secure SMS using ECC and RSA. In2013 13th International Conference on Intellient Systems Design and Applications 2013 Dec 8 (pp. 115-120). IEEE.

[57] Samarthrao KV, Rohokale VM. A hybrid meta-heuristic-based multi-objective feature selection with adaptive capsule network for automated email spam detection. International Journal of Intelligent Robotics and Applications 2022 Sep; 6(3):497–521.

[58] Shafi'i MA, Maryam S, Oluwafemi O, Ismaila I, John KA. Comparative analysis of classification algorithms for email spam detection.

[59] Du C, Huang L. Text classification research with attention-based recurrent neural networks. International Journal of Computers Communications & Control 2018 Feb 12;13(1):50–61.

[60] Lyubinets V, Boiko T, Nicholas D. Automated labeling of bugs and tickets using attention-based mechanisms in recurrent neural networks. In2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP) 2018 Aug 21 (pp. 271-275). IEEE.

[61] Zhang W. Spam filter through deep learning and information retrieval (Doctoral dissertation, Dissertation, Johns Hopkins University).

[62] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735–80.

[63] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86(11):2278–324.