Computer Systems
Science and Engineering

Tech Science Press

DOI: 10.32604/csse.2023.034456
Article

Automated Spam Review Detection Using Hybrid Deep Learning on Arabic Opinions

Ibrahim M. Alwayle[1], Badriyya B. Al-onazi[2], Mohamed K. Nour[3], Khaled M. Alalayah[1], Khadija M. Alaidarous[1], Ibrahim Abdulrab Ahmed[4], Amal S. Mehanna[5] and Abdelwahed Motwakel[6,*]

[1]Department of Computer Science, College of Science and Arts, Sharurah, Najran University, Saudi Arabia
[2]Department of Language Preparation, Arabic Language Teaching Institute, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia
[3]Department of Computer Sciences, College of Computing and Information System, Umm Al-Qura University, Saudi Arabia
[4]Computer Department, Applied College, Najran University, Najran, 66462, Saudi Arabia
[5]Department of Digital Media, Faculty of Computers and Information Technology, Future University in Egypt, New Cairo, 11845, Egypt
[6]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, AlKharj, Saudi Arabia
*Corresponding Author: Abdelwahed Motwakel. Email: a.ismaeil@psau.edu.sa
Received: 18 July 2022; Accepted: 04 November 2022

Abstract: Online reviews regarding purchasing services or products offered are the main source of users' opinions. To gain fame or profit, generally, spam reviews are written to demote or promote certain targeted products or services. This practice is called review spamming. During the last few years, various techniques have been recommended to solve the problem of spam reviews. Previous spam detection study focuses on English reviews, with a lesser interest in other languages. Spam review detection in Arabic online sources is an innovative topic despite the vast amount of data produced. Thus, this study develops an Automated Spam Review Detection using optimal Stacked Gated Recurrent Unit (SRD-OSGRU) on Arabic Opinion Text. The presented SRD-OSGRU model mainly intends to classify Arabic reviews into two classes: spam and truthful. Initially, the presented SRD-OSGRU model follows different levels of data preprocessing to convert the actual review data into a compatible format. Next, unigram and bigram feature extractors are utilized. The SGRU model is employed in this study to identify and classify Arabic spam reviews. Since the trial-and-error adjustment of hyperparameters is a tedious process, a white shark optimizer (WSO) is utilized, boosting the detection efficiency of the SGRU model. The experimental validation of the SRD-OSGRU model is assessed under two datasets, namely DOSC dataset. An extensive comparison study pointed out the enhanced performance of the SRD-OSGRU model over other recent approaches.

Keywords: Arabic text; spam reviews; machine learning; deep learning; white shark optimizer

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

As the Internet continues to grow in both factors, i.e., significance and size, the impact and quantity of online reviews are increasing. Reviews could influence individuals across a broad spectrum of industries, however specifically significant in the e-commerce domain [1], in which reviews and comments about services and products are often the most convenient one, if not the only, means for a purchaser to determine whether buy or not to buy the products. Online reviews are produced for several reasons [2]. Mostly, it is an effort to improve and enhance their businesses; service providers and online retailers might request their clients to give feedback regarding their experience with the services or products they bought and know if they are satisfied with the product or not [3]. Consumers might even feel inclined to review a service or product whenever they undergo a remarkably bad or good experience with those products or services [4]. While online reviews are helpful, trusting these reviews blindly becomes dangerous for buyers and sellers. Most of them glance at online reviews before making an order online, but reviews can be faked for gain or profit. Therefore any decision related to online reviews should be made carefully [5]. Additionally, proprietors may provide incentives to those who write good reviews regarding their goods or may pay to write negative reviews regarding services or products of competitors [6]. Such fake reviews were regarded as review spam and has a huge effect on online marketing because of it's the significance of reviews.

Review spam could also adversely affect businesses because of a loss of customer trust [7]. The problem has become very severe that grabs the attention of governments and mainstream media. Review spam is a damaging issue and pervasive; advancing techniques which will be helpful for consumers and entrepreneurs in distinguishing fake and truthful reviews somehow becomes difficult. Sentiment Analysis (SA) for Arabic texts is a research domain encompassing numerous challenging points that need proper dealing to achieve maximum performance [8]. Such challenges are the morphological complexities of the language, which state the demand for effectual preprocessing and feature representation, the demand for building accurate methods, and the necessity to detect and remove spam opinion texts. Morphological complexities surge because of the complicated nature of the Arabic language. For example, absence of standardization amongst the writing of similar words.

An Arabic word can be written in several formats which use many prefixes, suffixes, and affixes. Another difficulty is that Arabic spam opinion detection turns out to be one main task which is challenging that has a close relation to the analysis of opinions [9]. A spam opinion is a false or fake review. Spam opinion was generally written to destroy some reputation of a product by utilizing adverse opinions or promoting low-quality goods via positive opinions. Spam opinion detection has a huge effect on businesses due to the experiences of users being affected if the opinion given relating to a service or product encompasses a great amount of spammed opinion data. In addition, users do not buy or leverage this product or service again if they are cheated by such spam opinions [10]. Thus, devising a method that works well in detecting spam reviews in Arabic opinion texts becomes crucial, particularly since several works have already been developed mainly for the Arabic language. Though several works are available in the literature, it is necessary to focus on the hyperparameter tuning process.

This study develops an Automated Spam Review Detection using optimal Stacked Gated Recurrent Unit (SRD-OSGRU) on Arabic Opinion Text. The SRD-OSGRU model follows different levels of data preprocessing to convert the actual review data into a compatible format. Next, unigram and bigram feature extractors are utilized. The SGRU model is employed in this study to identify and classify Arabic spam reviews. Since the trial-and-error adjustment of hyperparameters is a tedious process, a white shark optimizer (WSO) is utilized, boosting the detection efficiency of the SGRU model. The experimental validation of the SRD-OSGRU model is assessed under two datasets, namely DOSC dataset.

2 Related Works

Saeed et al. [11] modelled a supervised learning technique for Arabic review sentiment classifications. This technique uses optimized compact features that rely on a well-representative feature set, combined with feature reduction algorithms that guarantee concurrent higher precision and space or time savings. The feature set involves a tripartite grouping of N-gram features, and negative or positive N-gram counts features gained after negation handling is taken into account. The presented technique analyses 2 distinct linear transformation techniques; latent Dirichlet allocation (LDA) as a supervised transformation algorithm and principal component analysis (PCA) as an unsupervised transforming approach. Ghourabi et al. [12] project a deep hybrid learning (DL) method to detect SMS spam messages. This detection algorithm depends on combining 2 DL long short-term memory (LSTM) and convolutional neural network (CNN). Its primary intention is to deal with a mixed text message written in English or Arabic.

Saeed et al. [13] contributed to this topic by introducing 4 distinct Arabic spam review detection techniques and paying greater attention to the evaluation and construction of an ensembling technique. The devised ensembling technique depends on compiling a rule-related classifier with machine learning (ML) approaches when using content-related features which rely on Negation handling and N-gram features. In [14], user and content attributes are studied to distinguish between illegitimate and legitimate users. After, uses these attributes with ML techniques for detecting spam on Twitter. It employs support vector machine (SVM) and Naïve Bayes (NB) classifier techniques for finding malicious content presented in the tweets. Bosaeed et al. [15] formulated a tool for detecting spam from outgoing SMS messages, even though the work is assigned to incoming and outgoing short messaging service (SMS) messages. To be Specific, it advances a mechanism with multiple ML-related classifiers constituted by employing 3 classifier techniques–NB, SVM, and NB Multinomial (NBM)- and 5 preprocessing and feature extracting techniques. El-Alfy et al. [16] see the effect of the imbalance ratio on the performance of Twitter spam detection by using multiple techniques of single and ensembling classifiers. Also, ensemble-related learning (Random Forest (RT) and Bagging) applied the SMOTE oversampling approach to improving detection performance, especially for classifiers sensitive to imbalanced data sets. [17] presents the ideology of implementing word-embedded related features with ML approaches for detecting Arabic spam tweets. Furthermore, the impact of the text domain of the corpus, which is collected for learning word embedding, was examined.

3 The Proposed Model

In this study, a new SRD-OSGRU technique has been developed for spam detection in Arabic reviews. Primarily, the presented SRD-OSGRU model follows different levels of data preprocessing to convert the actual review data into a compatible format. Then, unigram and bigram feature extractors are utilized. To classify the Arabic spam reviews into two classes, namely spam reviews and truthful reviews, the SGRU model is employed in this study. At last, the WSO is utilized as a hyperparameter optimizer to enhance the SGRU technique's detection efficiency. Fig. 1 depicts the block diagram of the SRD-OSGRU approach.

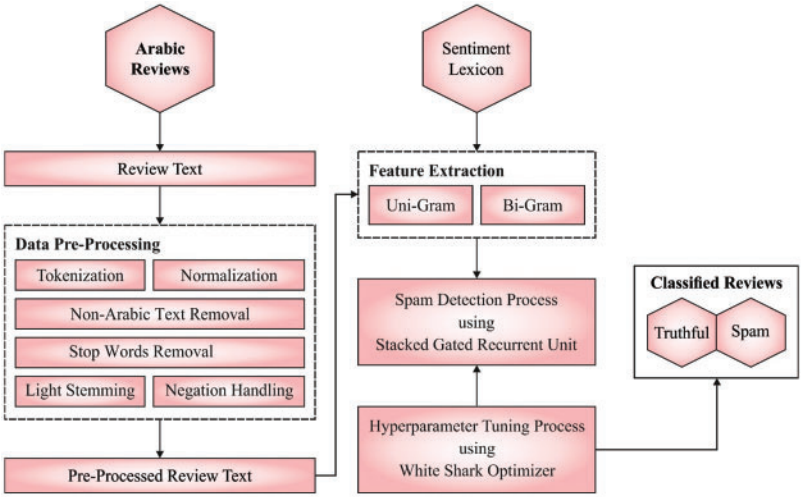


Figure 1: Block diagram of SRD-OSGRU approach

3.1 Data Preprocessing

At the introductory level, the presented SRD-OSGRU model follows different levels of data preprocessing. Preprocessing is accomplished to eliminate inappropriate parts of data beforehand feature extraction process. The preprocessing technique comprises five sequential phases: normalization, tokenization, light stemming, non-Arabic text, and stop words removal. Those phases are done initially on the reviews' text, which was significant to produce a preprocessed text ready for classification and feature extraction.

Tokenization: split the review's text into a series of tokens, whereas every token signifies a single word according to the whitespace character.

Normalization: convert the input data into a more generalized form.

Non-Arabic text removal: check each review's token to eliminate non-Arabic tokens in the review and to identify whether it is in normalized form.

Stop word removal: removing meaningless words frequently occurs in the review's text, which might reduce the index's space and improve the response time. Stop words are removed. (لقد ، فقط ، أمام ، كل ، في ، عن ، علي ، أو ، و ، كان ، من ، إلي ،)

Light Stemming: return the word to its original form. A basic stem is post-fixed or prefixed for expressing a grammatical syntax for non-Arabic languages. But it is hard to distinguish amongst some Arabic words afterwards stemming since those words have similar roots with totally different meanings.

Negation handling: considering Arabic negation words for accurate polarity classification. Fifty negators are used in the list of Arabic negation words (، ليس ما، ، لا ، لم ، لن ، مو ، مش etc...) constructed. Here, every extracted N-gram feature was checked to define whether or not the previous word was a negation word. The polarity of the N-gram feature can be retained if the word preceding is not a negation word, and it is reversed if the word preceding is a negation word.

3.2 Feature Extraction

Once the input data is preprocessed, the unigram and bigram feature extractors are utilized. Unigram refers to an arrangement of a single neighbouring component using a token of string elements as words, letters, or syllables. This arrangement signifies an n-gram for n = 1. The proportion of each unigram in a string of tokens is often applied for statistical text analysis in cryptography, computational linguistics, and speech recognition. Unigram assists in providing the conditional probability of a token by using the previous token while employing the relationship of conditional probability.

$$P\left(W_i|W_0 \ldots W_{i-1}\right) = P\left(W_i\right) \tag{1}$$

In Eq. (1), *P* refers to the conditional probability over the selected feature W.

A bigram signifies an arrangement of 2 neighbouring components using a string of tokens as words, letters, or syllables. A bigram is an n-gram for n = 2. The proportion of each bigram in a string is utilized for statistical text analysis in different fields like computational cryptography, linguistics, and speech recognition. Bigram assists in providing a conditional probability of a token by using the previous token while employing the relationship of conditional probability.

$$P(W_{n|}|W_{n-1}) \quad = \frac{P\left(W_{n-1}, W_n\right)}{P(W_{n-1})} \tag{2}$$

In Eq. (2), *P* refers to the conditional probability over the selected feature W.

3.3 Spam Detection Using SGRU Model

The SGRU model is utilized to classify Arabic spam reviews into two classes: spam reviews and truthful reviews [18]. The typical ML techniques manage time series problems; all the moments of instances assume distinct, independent arbitrary variables, and it can be provided as a regression method or NN to train. But, these techniques consider that the data at various moments were independent of each one, and its order from time is not assumed. A recurrent neural network (RNN) was presented using ML to capture this temporal correlation. The GRU is an improved RNN dependent upon LSTM. If the error signals propagate backwards with time from the typical RNN, the signals incline to vanish or blow up, and in both cases, this leads to the failure of networks for learning in data. The GRU not only maintains the capability for preventing the earlier revealed problems but also decreases the complexity of the infrastructure with no loss of the effectual learning capability.

The infrastructure of the GRU at all steps is the GRU cell. During this figure, the reset and update gates were fully connected (FC) layers with sigmoid activation that are utilized for controlling the memory. The preceding hidden layer (HL) maintains the memory, the reset gate controls for combining the input with memory for developing a candidate HL, and the update gate control for adding the candidate HL as HL. Lastly, the candidate HL, preceding HL, and resultant update gates establish the present HL and output. The GRU cell is demonstrated as follows:

$$\begin{aligned}
z_n &= \sigma\left(W^z x_n + U^z h_{n-1} + b^z\right),\\
r_n &= \sigma\left(W^r x_n + U^r h_{n-1} + b^r\right),\\
\tilde{h}_n &= \tanh(W x_n + U(r_n \odot h_{n-1}) + b),\\
h_n &= (1 - z_n)\,\tilde{h}_n + z_n \odot h_{n-1},
\end{aligned} \tag{3}$$

whereas h_{n-1} is the HL at $n - 1$ and x_n, z_n, r_n, $\tilde{h}_n$, and h_n is the input of GRU cell, resultant of the update gate, resultant of reset gate, candidate HL, and HL at n correspondingly. W and U imply the weighted matrices of the FC layer, and b denotes the bias vector. σ and tanh demonstrate the sigmoid and tanh activation function correspondingly. ⊙ refers to the element-wise product amongst 2 matrices of similar sizes. Fig. 2 illustrates the structure of the BiGRU technique. The present HL was linked to the next HL input for making the GRU. Several GRU cells are stacked beside the input-output directions to improve learning ability. The resultant GRU cells at every step are utilized as input for the next GRU cells at the equivalent step. Related to single-layer GRU, SGRU has several HLs that enhance the capability for learning time series.

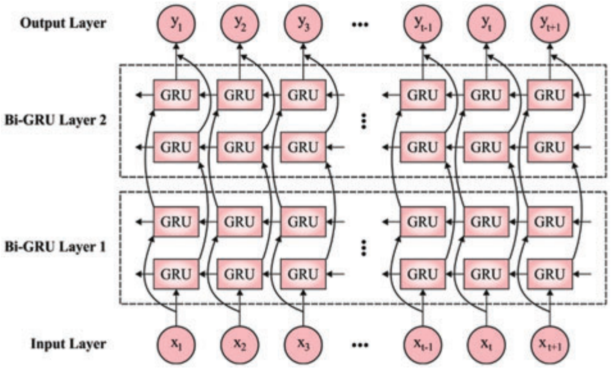


Figure 2: Framework of BiGRU

3.4 Hyperparameter Tuning Using WSO Algorithm

In this study, the WSO is utilized as a hyperparameter optimizer to enhance the detection efficiency of the SGRU model [19]. The mathematical model of the WSO algorithm involves the action of white sharks while hunting. This involves killing and tracking prey.

Initialization Process

A population of *n* WSO, in a *d* searching space, with the shark position representing a solution to this problem as follows.

$$w = \begin{bmatrix} w_1^1 & w_2^1 & \dots & \dots & w_d^1 \\ w_1^2 & w_2^2 & \dots & \dots & w_{\mathrm{d}}^2 \\ \dots & \dots & \dots & \dots & \dots \\ w_1^n & w_2^n & \dots & \dots & w_d^n \end{bmatrix} \tag{4}$$

In Eq. (4), *w* characterizes the position of each shark in the searching space, *and d* represents the number of selected parameters for a provided task.

Speed of Movement towards Prey

A white shark identifies a prey's position by hearing a pause in the wave as the prey moves.

$$u_{k+1}^i = \mu \left[u_k^i + p_1 \left(w_{gbest_k} - w_k^i \right) \times c_1 + p_2 \left(w_{best}^{v_k^i} - w_k^i \right) \times c_2 \right] \tag{5}$$

$i = 1, 2, \dots\dots, n =$ index of size *n*, and the novel speed vector of *i-th* shark is represented as $v^{i_{k+1}}$. v^i indicates the *i-th* index vector of a shark accomplishing the optimum position, as follows.

$$v = \lfloor n \times rand\,(1, n) \rfloor + 1 \tag{6}$$

In Eq. (6), rand $(1, n) =$ arbitrarily produced numbers within zero and one.

$$p_1 = p_{\max} + (p_{\max} - p_{\min}) \times e^{-(4k/k)^2} \tag{7}$$

$$p_2 = p_{\min} + (p_{\max} - p_{\min}) \times e^{-(4k/k)^2} \tag{8}$$

Now k = current, K = maximal iteration, $p_{\min}$ and $p_{\max}$ denote the starting and subordinate velocities for the movement of the white shark. $p_{\min}$ and $p_{\max}$ values are found to be 0.5 and 1.5 after a thorough examination,

$$\mu = \frac{2}{|2 - \tau - \sqrt{\tau^2 - 4\tau}|} \tag{9}$$

Here, τ indicates the accelerating factor that is 4.125.

Movement Towards Optimal Prey

The location update system determines the behaviour of white sharks once they move toward the prey

$$w^i_{k+1} = \begin{cases} w^i_k. \rightarrow \oplus w_0 + u.a + l.b, rand < mv \\ w^i_k + u^i_k/f, rand \geq mv \end{cases} \tag{10}$$

From the equation, a and b refer to binary vectors.

$$a = sgn\left(w^i_k - u\right) > 0 \tag{11}$$

$$b = sgn\left(w^i_k - 1\right) < 0 \tag{12}$$

$$w_0 = \oplus\,(a, b) \tag{13}$$

Here, $\oplus$ indicates the outcome of a bitwise XOR operation. The frequency of white shark's wavy movement and the multitude of times they attack the prey can be defined as follows

$$f = f_{\min} + \frac{f_{\max} - f_{\min}}{f_{\max} + f_{\min}} \tag{14}$$

$$mv = \frac{1}{(a_0 + e^{(k/2-k)/a_1})} \tag{15}$$

In Eq. (15), a_0 and a_1 represent the constants that regulate exploration and exploitation.

Movement Towards Optimal Shark

Sharks keep their position in front of the best one nearby to the prey, expressed in the following equation.

$$w'^i_{k=1} = w_{gbestk} + r_1 \overrightarrow{D_w}\, sgn\,(r_2 - 0.5)\, r_3 < S_s \tag{16}$$

$w'^i{k+1}$ = Upgraded shark position, $sgn(r_2 - 0.5)$ return 1 or -1 to adapt the search path, r_1, r_2, and $r_3 = rand$. The random value lies between zero and one, D_w = length for the target and shark. S_s indicates a parameter that reflects the power of white sharks as follows.

$$\overrightarrow{D_w} = \left|rand \times \left(w_{gbest} - w^i_k\right)\right| \tag{17}$$

$$S_s = \left|1 - e^{\left(-a_2 \times \frac{k}{k}\right)}\right| \tag{18}$$

Now, a_2 indicates a position factor utilized for control exploration and exploitation.

Algorithm 1: Pseudocode of WSO Algorithm

```
Initializing the parameter of the problem
Initializing the parameter of WSO
Arbitrarily produce the primary position of WSO
Initializing velocity of the early population
Assess the location of the early population
while ($k < K$) do
Upgrade the variables $v$, $p_1$, $p_2$, $\mu$, $a$, $b$, $w_0$,$f$, $m_v$ and $S_s$ using Eqs. (6)–(9), (11)–(18), and (18), correspondingly.
for $i = 1$ to $n$ do
$v^{i_{k+1}} = \mu[v^{i_k} + p_1(w_{gbestk} - w^{i_k}) \times c_1 + p_2(w'^{vk_{best}} - w^{i_k}) \times c_2]$
end for
for $i = 1$ to $n$ do
        if $rand < mv$ then
            $w^{i_{k+1}} = w^{i_k} \cdot - \oplus w_0 + u \cdot a + l \cdot b$
         else
            $w^{i_{k+1}} = w^{i_k} + v^{i_k}/f$
        end if
    end for
for $i = 1$ to $n$ do
        if $rand \leq S_s$ then
            $\overrightarrow{D}_w = |rand \times (w_{gbest} - w^i_k)|$
            if $i == 1$ then
            $w^i_{k+1} = w_{gbestk} + r_1\overrightarrow{D}_w \text{ sgn } (r_2 - 0.5)$
            else
            $w'^i_{k=1} = w_{gbestk} + r_1\overrightarrow{D}_w \text{ sgn } (r_2 - 0.5)$
            $w^i_{k=1} = \frac{w^i_k + w'^i_{k+1}}{2 \times rand}$
                  endif
         end if
    end for
Alter the location of the white shark proceeds beyond the boundary
Assess and upgrade the new position
$k = k + 1$
end while
Return the optimum solution
```

The WSO method derives a fitness function (FF) from improving classifier outcomes. It sets a positive number for representing the superior outcome of the candidate solutions. In this article, the reduction of the classifier error rate can be regarded as the FF, as provided in Eq. (19). The finest solution comprises the least error rate, and the poor solution reaches a maximal error rate.

$$fitness\,(x_i) = Classifier\ Error\ Rate\,(x_i)$$
$$= \frac{number\ of\ misclassified\ samples}{Total\ number\ of\ samples} * 100 \qquad (19)$$

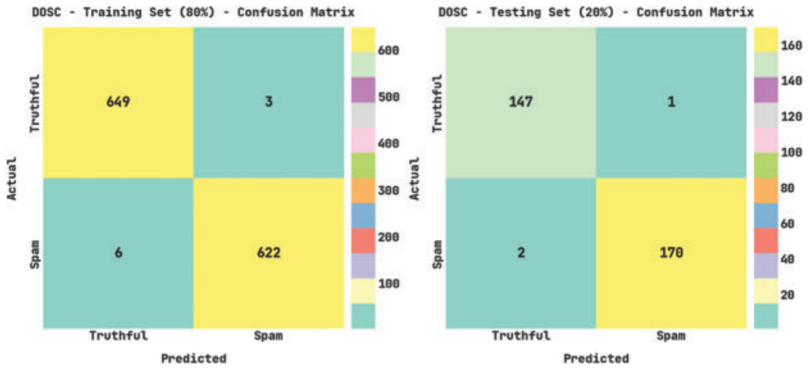
4 Results and Discussion

This section tests the SRD-OSGRU model's experimental validation using a benchmark dataset [20]. The proposed model is simulated using Python 3.6.5 tool on PC i5-8600k, GeForce 1050Ti 4 GB, 16 GB RAM, 250 GB SSD, and 1TB HDD. The parameter settings are given as follows: learning rate: 0.01, dropout: 0.5, batch size: 5, epoch count: 50, and activation: ReLU. The DOSC dataset includes 1600 reviews with the inclusion of 800 truthful reviews and 800 spam reviews. Table 1 illustrates the detailed description of the DOSC dataset.

Table 1: Details of DOSC dataset

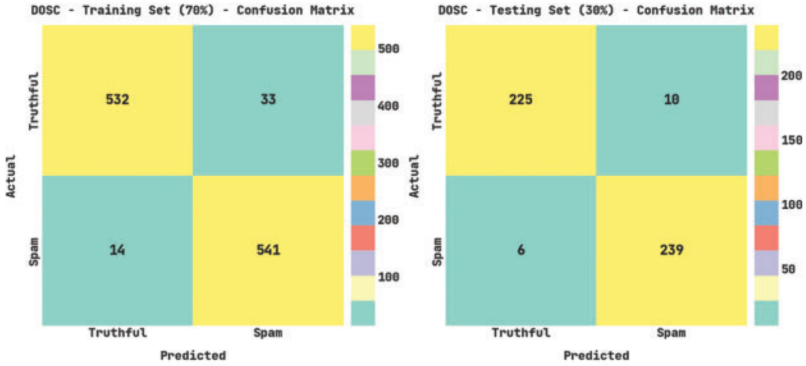
Deceptive Opinion Spam Corpus (DOSC)	
Class	No. of reviews
Truthful	800
Spam	800
Total No. of reviews	1600

Fig. 3 illustrates the confusion matrices produced by the SRD-OSGRU model on the DOSC dataset under varying training (TR) and testing (TS) data. With 80% of TR data, the SRD-OSGRU technique has recognized 649 samples into the truthful class and 622 samples under the spam class. Moreover, with 20% of TR data, the SRD-OSGRU method has recognized 147 samples into the truthful class and 170 samples under the spam class. Also, with 70% of TR data, the SRD-OSGRU algorithm has recognized 532 samples into the truthful class and 541 samples under the spam class. In the meantime, with 30% of TS data, the SRD-OSGRU approach has recognized 225 samples into the truthful class and 239 samples under the spam class.



(a) (b)

Figure 3: (Continued)



(c) (d)

Figure 3: Confusion matrices of SRD-OSGRU approach under DOSC dataset (a) 80% of TR data, (b) 20% of TS data, (c) 70% of TR data, and (d) 30% of TS data

Table 2 and Fig. 4 highlight the spam review detection and classification outcomes on the DOSC dataset. The obtained values implied that the SRD-OSGRU model had shown enhanced results under all classes. For instance, with 80% of TR data, the SRD-OSGRU algorithm has attained average $accu_y$, $prec_n$, $reca_l$, $spec_y$, and F_{score} of 99.30%, 99.30%, 99.29%, 99.29%, and 99.30%, respectively. Then, with 20% of TR data, the SRD-OSGRU approach has attained average $accu_y$, $prec_n$, $reca_l$, $spec_y$, and F_{score} of 99.06%, 99.04%, 99.08%, 99.08%, and 99.06% correspondingly. Then, with 70% of TR data, the SRD-OSGRU algorithm has acquired average $accu_y$, $prec_n$, $reca_l$, $spec_y$, and F_{score} of 95.80%, 95.84%, 95.82%, 95.82%, and 95.80% correspondingly. Finally, with 30% of TR data, the SRD-OSGRU technique has acquired average $accu_y$, $prec_n$, $reca_l$, $spec_y$, and F_{score} of 96.67%, 96.69%, 96.65%, 99.65%, and 96.66% correspondingly.

Table 2: Result analysis of the SRD-OSGRU model with different measures under DOSC dataset

Labels	Accuracy	Precision	Recall	Specificity	F-score
Training set (80%)					
Truthful	99.30	99.08	99.54	99.04	99.31
Spam	99.30	99.52	99.04	99.54	99.28
Average	99.30	99.30	99.29	99.29	99.30
Testing set (20%)					
Truthful	99.06	98.66	99.32	98.84	98.99
Spam	99.06	99.42	98.84	99.32	99.13
Average	99.06	99.04	99.08	99.08	99.06
Training set (70%)					
Truthful	95.80	97.44	94.16	97.48	95.77
Spam	95.80	94.25	97.48	94.16	95.84
Average	95.80	95.84	95.82	95.82	95.80

(Continued)

Table 2: Continued

Labels	Accuracy	Precision	Recall	Specificity	F-score
Testing set (30%)					
Truthful	96.67	97.40	95.74	97.55	96.57
Spam	96.67	95.98	97.55	95.74	96.76
Average	96.67	96.69	96.65	96.65	96.66

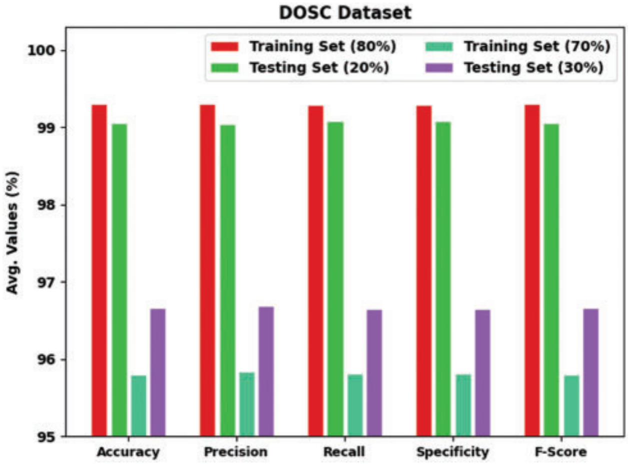


Figure 4: Average analysis of the SRD-OSGRU approach under DOSC dataset

The training accuracy (TA) and validation accuracy (VA) acquired by the SRD-OSGRU method on the DOSC dataset is illustrated in Fig. 5. The experimental outcome denoted the SRD-OSGRU technique maximal values of TA and VA. In specific, the VA is greater than TA.

The training loss (TL) and validation loss (VL) attained by the SRD-OSGRU approach on the DOSC dataset are shown in Fig. 6. The experimental outcome is implicit in the SRD-OSGRU algorithm accomplished least values of TL and VL. Particularly, the VL is lesser than TL.

Table 3 and Figs. 7 and 8 offer a detailed comparative inspection of the SRD-OSGRU model with recent models on DOSC dataset [13]. The experimental values inferred that the SRD-OSGRU model had improved performance with maximum classification results. Based on $accu_y$, the SRD-OSGRU model has depicted a higher $accu_y$ of 99.30%. In contrast, the stacking ensemble, RB-Boosting, RB-RF, RB-NN, RB-Bagging, RB-KNN, and RB-K-means models have exhibited lower $accu_y$ of 94.87%, 85.36%, 85.47%, 86.07%, 85.57%, 86.36%, and 95.51% respectively. Temporarily, based on $prec_n$, the SRD-OSGRU model has depicted a higher $prec_n$ of 99.30%. In contrast, the stacking ensemble, RB-Boosting, RB-RF, RB-NN, RB-Bagging, RB-KNN, and RB-K-means algorithms have displayed lower $prec_n$ of 98.53%, 97.54%, 97.85%, 99.19%, 97.63%, 97.86%, and 99.09% correspondingly; Finally, based on $reca_l$, the SRD-OSGRU algorithm has depicted higher $reca_l$ of 99.29%. In contrast, the stacking ensemble, RB-Boosting, RB-RF, RB-NN, RB-Bagging, RB-KNN, and RB-K-means models have exhibited lower $reca_l$ of 91.78%, 72.21%, 72.90%, 72.20%, 73.07%, 73.93%, and 91.47% correspondingly. Besides, based on $spec_y$, the SRD-OSGRU technique has depicted a higher $spec_y$ of 99.29%. In contrast, the stacking ensemble, RB-Boosting, RB-RF, RB-NN, RB-Bagging, RB-KNN, and RB-K-means models have exhibited lower $spec_y$ of 98.33%, 97.96%, 98.63%, 99.05%, 98.44%, 98.87%, and 98.37% correspondingly. Finally, based on F_{score}, the SRD-OSGRU method has depicted a higher F_{score} of 99.30%, whereas the stacking ensemble, RB-Boosting, RB-RF, RB-NN, RB-Bagging, RB-KNN, and RB-K-means models have displayed lower F_{score} of 98.60%, 82.82%, 83.34%, 83.15%, 83.49%, 84.62%, and 95.28% correspondingly.

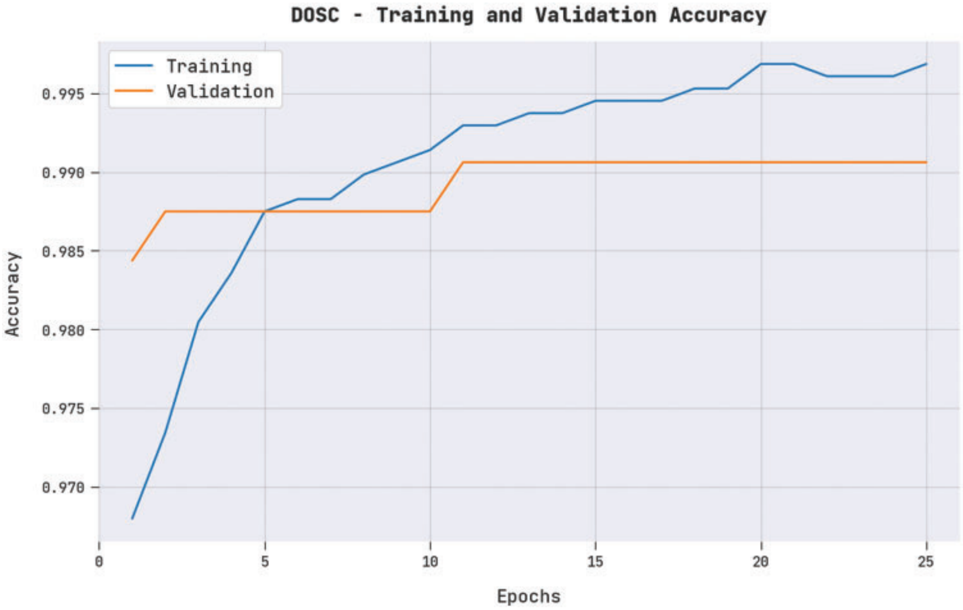


Figure 5: TA and VA analysis of SRD-OSGRU approach under DOSC dataset

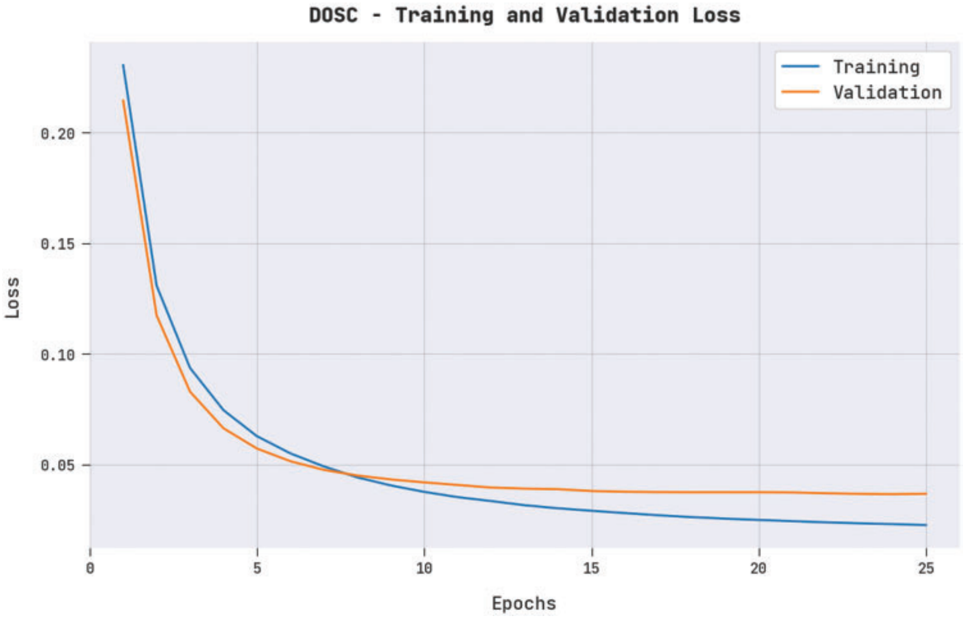


Figure 6: TL and VL analysis of SRD-OSGRU approach under DOSC dataset

Table 3: Comparative analysis of SRD-OSGRU approach with recent algorithms on DOSC dataset

Methods	Accuracy	Precision	Recall	Specificity	F1 score
SRD-OSGRU	99.30	99.30	99.29	99.29	99.30
Stacking ensemble	94.87	98.53	91.78	98.33	98.60
RB-Boosting	85.36	97.54	72.21	97.96	82.82
RB-RF model	85.47	97.85	72.90	98.63	83.34
RB-NN model	86.07	99.19	72.20	99.05	83.15
RB-Bagging	85.57	97.63	73.07	98.44	83.49
RB-KNN model	86.39	97.86	73.93	98.87	84.62
RB-K-Means	95.51	99.06	91.47	98.37	95.28

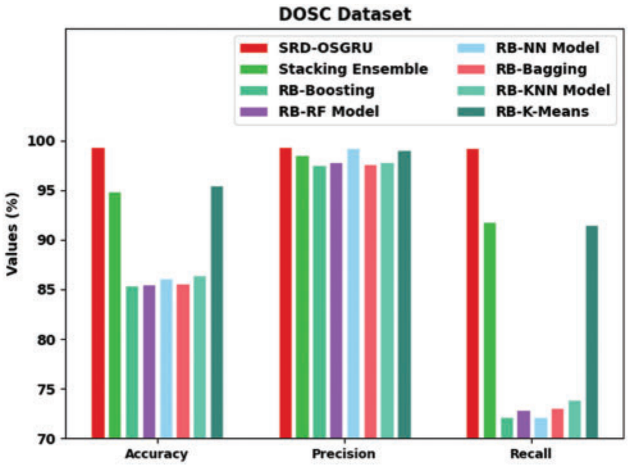


Figure 7: $Accu_y$, $prec_n$, and $reca_l$ analysis of SRD-OSGRU approach with recent algorithms DOSC dataset

Meanwhile, based on $reca_l$, the SRD-OSGRU approach has depicted a higher $reca_l$ of 99%. In contrast, the stacking ensemble, RB-NB, RB-SVM, RB-RF, RB-LOR, RB-Bagging, and RB-NN models have exhibited lower $reca_l$ of 98.51%, 96.82%, 97.68%, 98.39%, 98.22%, 97.87%, and 98.15% correspondingly. Eventually, based on $spec_y$, the SRD-OSGRU algorithm has depicted a higher $spec_y$ of 99%, whereas the stacking ensemble, RB-NB, RB-SVM, RB-RF, RB-LOR, RB-Bagging, and RB-NN models have exhibited lower $spec_y$ of 98.60%, 98.54%, 98.39%, 97.98%, 98.43%, 97.99%, and 98.53% correspondingly. Next, based on F_{score}, the SRD-OSGRU model has depicted a higher F_{score} of 99%, whereas the stacking ensemble, RB-NB, RB-SVM, RB-RF, RB-LOR, RB-Bagging, and RB-NN models have exhibited lower F_{score} of 98.40%, 97.59%, 98.17%, 98.18%, 98.32%, 98.72%, and 98.55% correspondingly. From the detailed results and discussion, it is apparent that the SRD-OSGRU model has demonstrated maximum performance in spam detection in Arabic text.

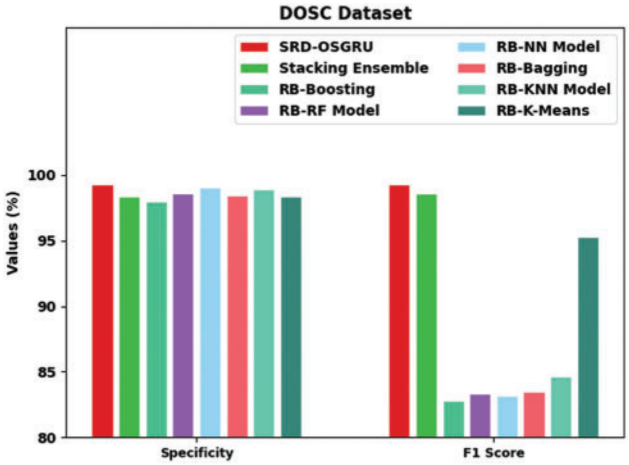


Figure 8: $Spec_y$ and $F1_{score}$ analysis of SRD-OSGRU approach with recent algorithms DOSC dataset

5 Conclusions and Future Directions

This study developed a new SRD-OSGRU algorithm for identifying and classifying Arabic spam reviews. Primarily, the presented SRD-OSGRU model follows different levels of data preprocessing to convert the actual review data into a compatible format. Then, unigram and bigram feature extractors are utilized. To classify the Arabic spam reviews into two classes, namely spam reviews and truthful reviews, the SGRU model is employed in this study. At last, the WSO is utilized as a hyperparameter optimizer to enhance the SGRU method's detection efficiency. The experimental validation of the SRD-OSGRU model is assessed under two datasets, namely the DOSC dataset. An extensive comparison study pointed out the enhanced performance of the SRD-OSGRU model over other recent approaches. In the future, an ensemble of DL-based fusion models with hybrid metaheuristics can be designed to boost the detection efficiency of the SRD-OSGRU method.

Funding Statement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R263), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: 22UQU4310373DSR58. The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work under the Research Groups Funding program grant code (NU/RG/SERC/11/7).

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

References

[1] H. Najadat, M. A. Alzubaidi and I. Qarqaz, "Detecting Arabic spam reviews in social networks based on classification algorithms," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 1, pp. 1–13, 2022.

[2] I. Amin and M. K. Dubey, "An overview of soft computing techniques on Review Spam Detection," in *2021 2nd Int. Conf. on Intelligent Engineering and Management (ICIEM)*, London, United Kingdom, pp. 91–96, 2021.

[3] M. Z. Asghar, A. Ullah, S. Ahmad and A. Khan, “Opinion spam detection framework using hybrid classification scheme,” *Soft Computing*, vol. 24, no. 5, pp. 3475–3498, 2020.

[4] I. Amin and M. K. Dubey, “Hybrid ensemble and soft computing approaches for review spam detection on different spam datasets,” *Materials Today: Proceedings*, vol. 62, pp. 4779–4787, 2022.

[5] A. S. Alhassun and M. A. Rassam, “A combined text-based and metadata-based deep-learning framework for the detection of spam accounts on the social media platform twitter,” *Processes*, vol. 10, no. 3, pp. 439, 2022.

[6] O. El Kouari, H. Benaboud and S. Lazaar, “Using machine learning to deal with Phishing and Spam Detection: An overview,” in *Proc. of the 3rd Int. Conf. on Networking, Information Systems & Security*, Marrakech Morocco, pp. 1–7, 2020.

[7] M. E. Basiri, N. Safarian and K. H., “Farsani, a supervised framework for review spam detection in the persian language,” in *2019 5th Int. Conf. on Web Research (ICWR)*, Tehran, Iran, pp. 203–207, 2019.

[8] S. K. Maurya, D. Singh and A. K. Maurya, “Deceptive opinion spam detection approaches: A literature survey,” *Applied Intelligence*, 2022. https://doi.org/10.1007/s10489-022-03427-1

[9] N. H. Imam, V. G. Vassilakis and D. Kolovos, “An empirical analysis of health-related campaigns on twitter Arabic hashtags,” in *2022 7th Int. Conf. on Data Science and Machine Learning Applications (CDMA)*, Riyadh, Saudi Arabia, pp. 29–41, 2022.

[10] A. Ziani, N. Azizi, D. Schwab, D. Zenakhra, M. Aldwairi *et al.,* “Deceptive opinions detection using new proposed arabic semantic features,” *Procedia Computer Science*, vol. 189, no. 6, pp. 29–36, 2021.

[11] R. M. K. Saeed, S. Rady and T. F. Gharib, “Optimizing sentiment classification for Arabic opinion texts,” *Cognitive Computation*, vol. 13, no. 1, pp. 164–178, 2021.

[12] A. Ghourabi, M. A. Mahmood and Q. M. Alzubi, “A hybrid cnn-lstm model for SMS spam detection in Arabic and English messages,” *Future Internet*, vol. 12, no. 9, pp. 156, 2020.

[13] R. M. K. Saeed, S. Rady and T. F. Gharib, “An ensemble approach for spam detection in Arabic opinion texts,” *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 1, pp. 1407–1416, 2022.

[14] D. Alorini and D. B. Rawat, “Automatic spam detection on gulf dialectical Arabic tweets,” in *2019 Int. Conf. on Computing, Networking and Communications (ICNC)*, Honolulu, HI, USA, pp. 448–452, 2019.

[15] S. Bosaeed, I. Katib and R. Mehmood, “A fog-augmented machine learning based SMS spam detection and classification system,” in *2020 Fifth Int. Conf. on Fog and Mobile Edge Computing (FMEC)*, Paris, France, pp. 325–330, 2020.

[16] E. S. M. El-Alfy and S. Al-Azani, “Statistical comparison of opinion spam detectors in social media with imbalanced datasets,” in *Int. Symp. on Security in Computing and Communication, Communications in Computer and Information Science Book Series*, Singapore, Springer, vol. 969, pp. 157–167, 2018.

[17] S. Al-Azani and E. -S. M. El-Alfy, “Detection of Arabic spam tweets using a word embedding and machine learning,” in *2018 Int. Conf. on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, Sakhier, Bahrain, pp. 1–5, 2018.

[18] H. M. Lynn, S. B. Pan and P. Kim, “A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks,” *IEEE Access*, vol. 7, pp. 145395–145405, 2019.

[19] M. Braik, A. Hammouri, J. Atwan, M. A. Al-Betar and M. A. Awadallah, “White shark optimizer: A novel bio-inspired meta-heuristic algorithm for global optimization problems,” *Knowledge-Based Systems*, vol. 243, no. 7, pp. 108457, 2022.

[20] Dataset: Deceptive Opinion Spam Corpus | Kaggle, 2022. https://www.kaggle.com/datasets/rtatman/deceptive-opinion-spam-corpus