# Explainable Artificial Intelligence and Natural Language Processing for Unraveling Deceptive Contents

**Nadezda Pospelova[1], Aiziryak Tarasova[2], Natalya Subbotina[3], Natalya Koroleva[4], Nilufar Raimova[5], E. Laxmi Lydia[6]**

[1]Candidate of Philological Sciences, Associate Professor of Department of English Philology and Intercultural Communication, Kazan Federal University, Elabuga Institute of KFU, Elabuga, Russia
[2]Candidate of Philological Sciences, Senior Lecturer of Department of English Philology and Intercultural Communication, Kazan Federal University, Elabuga Institute of KFU, Elabuga, Russia.
[3]Senior Lecturer of Department of English Philology and Intercultural Communication, Kazan Federal University, Elabuga Institute of KFU, Elabuga, Russia;
[4]Candidate of Pedagogical Sciences, Associate Professor of Department of English Philology and Intercultural Communication, Kazan Federal University, Elabuga Institute of KFU, Elabuga, Russia;
[5]Senior Lecturer of Primary Education Methodology Department, Urgench State Pedagogical Institute, Urgench, Uzbekistan;
[6]Department of Computer Science and Engineering, GMR Institute of Technology, Andhra Pradesh, Rajam 532127, India;
Emails: pospelova.n.v@mail.ru; tarasova.a.n@inbox.ru; subbotina.n.s@inbox.ru; koroleva.n.e@list.ru; raimova.n.a@mail.ru; elaxmi2002@yahoo.com

**Abstract**

Deceptive content recognition in social media employing artificial intelligence (AI) includes the use of sophisticated techniques and machine learning (ML) methods to recognize deceptive or wrong data shared on numerous platforms. AI methods analyse textual as well as multimedia content, investigative patterns, linguistic cues, and contextual info to flag latent cases of deception. As a result of the use of natural language processing (NLP) and computer vision (CV), these systems identify subtle nuances, misrepresentation strategies, and anomalies in user-generated content. This active technique permits social media platforms, organizations, and consumers to recognize and diminish the spread of deceptive content, donates to a more reliable online atmosphere, and aids in fighting tasks modelled by misinformation and false news. This study offers a novel sine cosine algorithm with deep learning-based deceptive content detection on social media (SCADL-DCDSM) technique. The SCADL-DCDSM technique incorporates the ensemble learning process with a hyperparameter tuning strategy for classifying the sentiments. Primarily, the SCADL-DCDSM technique pre-processes the input data to change the input data into a valuable format. Moreover, the SCADL-DCDSM algorithm follows the BERT model for the word embedding process. Moreover, the SCADL-DCDSM technique involves an ensemble of three models for sentiment classification such as long short-term memory (LSTM), extreme learning machine (ELM), and attention-based recurrent neural network (ARNN). Finally, SCA can be executed for better hyperparameter choice of the DL models. The SCADL-DCDSM system integrates the explainable artificial intelligence (XAI) system LIME has been employed for a comprehensive explainability and understanding of the black-box process, enhancing correct deceptive content recognition. The simulation result analysis of the SCADL-DCDSM algorithm has been examined on a benchmark database. The simulation outcome illustrated that the SCADL-DCDSM methodology achieves optimum solution than other approaches in terms of different measures.

**Keywords:** Social Media; Word Embedding; Explainable Artificial Intelligence; BERT; Natural Language Processing



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024

## 1. Introduction

Online social network interconnects the public, permitting videos, images, personal news, and alternative content that can be simply shared with family, friends, and someone else who attentions to read it [1]. The achievements of every organization's or individual's content depends on who selects to read it directly and who selects for re-sharking it. In this context, social media platforms are democratized information content [2]. The method has eliminated the filter of news media groups—permitting content to flow unconstrained (in a few conditions, un-fact-checked and unedited) from writer to reader promptly. This person-to-person communication ability enables social development. It enables members of the public banding collected to need the righting of wrongs [3]. Actions namely the "Arab Spring" uprisings have exhibited the robustness of social media coordination (still, some have reduced its function or developed social media employ have been products of objections rather than a cause for them). However, this could be offered as a platform for those who try to disseminate misinformation [4]. Any misinformation could be recognized as benign causes, namely as various standpoints under actions or posting users themselves believing inaccurate causes. In alternative conditions, organizations and individuals post content considering that it is incorrect and to handle readers. One popularly states that disinformation transfers false data deliberately broadcasted for deception [5]. Alternatively, the unplanned spread of misleading and wrong data has been named as misinformation. Deception or false information can be types of flavours such as Misinformation, Click-Baits, Fake or False News, Satire, Junk News, Hoaxes, Rumors, Disinformation, Propaganda, and so on.

An important number of technical approaches are designed to automatically exceed the rumors online by implementing machine learning (ML) methods [6]. The fake news identification technique employs different features to combat fake news in the social network, which contains news content, user data, and the proliferation of each social post. The ML-based traditional techniques for fake news identification leverage hand-crafted features as well as accept the diverse functions of real posts and retweets in the generation of data and distributing method that makes the present identification ineffective [7]. Similarly, as the scale of data developments, the ML technique appears to fall behind the deep learning (DL) technique. The DL is an endeavour for learning higher-level feature representations across the hidden layers (HLs) of large-scale data to offer accurate outcomes. This representation has a different part of the DL method and is a major stage ahead of ML approaches [8]. In recent times, the increase in analyzing fake news under social media platforms could be aimed at resolving serious issues by employing the DL gathering with web services and developing databases to determine the deceiving contents [9]. Thus, major prior research work deeply depends on the DL method. But, even with the DL method, this can be extremely challenging to define if the news data has been fake or real when received with fake news identification. For instance, when people accept a decision, they cross-refer these user comments that the reliance on the commentaries supports for recognizing fake news [10]. Besides, the user comment analysis supports individuals in obtaining accurate data.

This study develops a novel sine cosine algorithm with DL-based deceptive content detection on social media (SCADL-DCDSM) technique. The SCADL-DCDSM technique incorporates the ensemble learning process with hyperparameter tuning strategy for classifying the sentiments. Primarily, the SCADL-DCDSM technique pre-processes the input data to change the input data into a suitable format. Moreover, the SCADL-DCDSM technique involves an ensemble of three models for sentiment classification such as long short-term memory (LSTM), extreme learning machine (ELM), and attention-based recurrent neural network (ARNN). Finally, SCA can be executed for better hyperparameter choice of the DL models. The SCADL-DCDSM system integrates the explainable artificial intelligence (XAI) system LIME has been employed for a comprehensive explainability and understanding of the black-box process, enhancing correct deceptive content recognition. The simulation result analysis of the SCADL-DCDSM methodology has been examined on a benchmark database.

## 2. Related Work

Hamed et al. [11] examined a Bidirectional LSTM (Bi-LSTM)-based algorithm that has trained under an Islamic database (RIDI) to be gathered and labeled by 2 distinct specified groups. Furthermore, employing a pre-trained word-embedding system could be generated out of vocabulary due to its handled with a specific field. To overcome this problem, the method was re-trained the pre-trained Glove architecture under Islamic documents employing the Mittens technique. In [12], a hybrid DL algorithm was introduced that employs convolutional neural network (CNN) and LSTM systems for identifying COVID19 false information. This developed method includes a few layers namely convolutional layer, flatten layer, embedding layer, pooling layer, dense layer, output layer, and LSTM layer. For validation outcomes, 3 COVID-19 fake news databases have been employed for an estimated 6 ML techniques, a 2 DL methods. Fouad et al. [13] designed a framework to identify fake news by employing only textual features. ML and DL approaches should be utilized. The DL methods could be implemented dependent on LSTM, CNN, CNN+LSTM, CNN + Bi-LSTM, and Bi-LSTM.

Sadiq et al. [14] introduced a modest DL method in integration with word embedding employing an openly accessible Tweepfake database. A standard CNN method was developed, leveraging FastText word embedding to accept the function of recognizing deepfake tweets. This analysis implemented numerous ML techniques as



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024

baseline approaches for comparison. These baseline approaches employed different features, comprising FastText subword embeddings, FastText, Term Frequency-Inverse, Term Frequency, and Document Frequency. In [15], the authors built an effectual technique for identifying and categorizing toxicity in media platforms in user-generated content by applying the BERT method. The BERT pre-trained method and 3 of its variations must be adjusted under an identified label toxic comment database, called the Kaggle open dataset. Kaliyar et al. [16] introduced a multichannel DCNN system with diverse kernel sizes and filters as an AI technology. Numerous embedding of similar dimensions with a variety of kernel sizes precisely permit the news articles that be simultaneously processed in diverse resolutions of several n-grams.

Prakash and Kumar [17] projected a DL-based multimodal framework, which employs Attention Mechanism for fake news identification. The analysis employed the benchmark Weibo database gathered from the Weibo micro-blogging Website, comprising visual and textual information. LSTM has been employed to handle textual data as well as diverse pre-trained DL methods for handling visual data. The attention mechanism combines visual and textual features, and lastly, the softmax layer has been employed for classification. Kumar and Sachdeva [18] aimed at a hybrid architecture, BiGRU-Attention-CapsNet (Bi-GAC), which advantages by learning spatial location data and sequential semantic representations employing a BiGRU with self-attention and then, CapsNet in cyberbullying identification at the textual content of social networking platform.

## 3. The Proposed Method

In this study, we introduce a new SCADL-DCDSM methodology. The SCADL-DCDSM technique incorporates the ensemble learning process with hyperparameter tuning strategy for classifying the sentiments. Fig. 1 demonstrates the entire procedure of the presented SCADL-DCDSM algorithm.

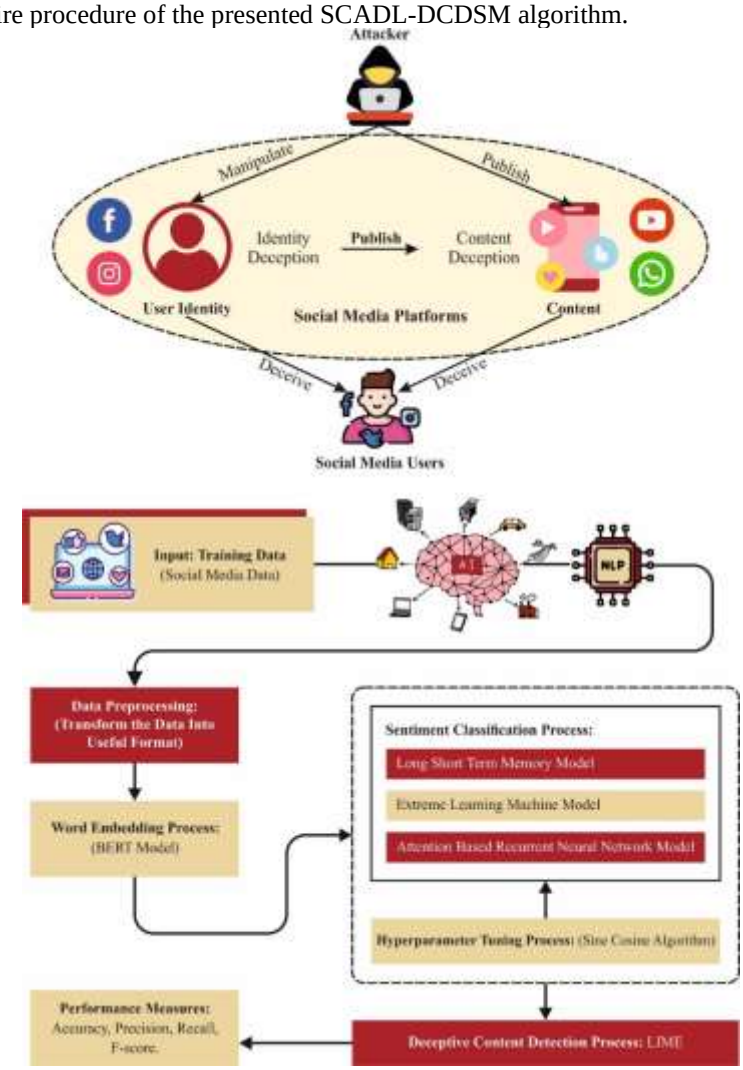


Figure 1: Overall process of SCADL-DCDSM technique



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024

### A. Data Pre-processing

Pre-processing for deceptive content recognition in social media comprises numerous main steps to increase the accurateness of the following analysis. This procedure classically starts with text lowercasing to safeguard evenness and extraction of irrelevant characters that include special symbols and numbers. Tokenization breaks down text into separate units for study, while stopword elimination removes common words that do not donate considerably to deception recognition. Handling user-specific elements like mentions and usernames is vital to concentrate on content's deceptive nature. Emoji handling and emoticons along with removal of hyperlinks, aid in updating text for analysis. Spell checking guarantees correct interpretation, and stemming or lemmatization decreases words to their root forms. Lastly, addressing abbreviations, slang, and the influence of negations further refines data, optimizing it for deceptive content recognition by employing AI and NLP models.

### B. BERT Model

One of the famous and commonly employed pre-trained language techniques that comprises a modifier is the BERT technique [19]. It pre-trains a great corpus. With the usage of bi-directional transformers, and then regulated to perform specific challenges. BERT is a pre-processing method for NLP that employs NNs. Word2vec is not capable of regulating how it signifies context-dependent words. By contrast, Google prepared its source code as openly accessible. Consumers only want to make slight changes to pre-trained BERT techniques depending on the particulars of the present work to save both money and time. Employing the trained BERT method, it can be probable to use it a huge range of downstream challenges from the NLP. The top HL of the transformer system signifies word vector matrix $S_D$ of sentence $S$.

$$S_D = [X_0, X_1 \cdots X_n, X_{n+1}] \quad (1)$$

The submatrix utilized to signify desired words $D_r$.

$$D_r = [X_1, X_1 + 1, \cdots X_{1+m-1}] \quad (2)$$

$D_r \in R^{ml}$, where $m$ denotes the preferred length. The target vectors can be executed to a max-pooling process that picks the most vital features in every word from the target at every dimensional.

$$y = \max\{D_r, L = 0\}, V \in R^{1*L} \quad (3)$$

### C. Ensemble Learning Based Classification

The easy technique for integrating forecasts of manifold methods is the averaging approach [20]. It commonly utilizes the ensemble method in which all the models are trained distinctly, and the averaging method linearly combines every prediction of models by averaging them to provide the last prediction. This model can be used without the need for additional training on vast numbers of separate predictions. Typically, voting has been the standard method for averaging the prediction of baseline classifiers. The last prediction outcomes are frequently defined by a majority vote on predictive of numerous classifiers stated as hard voting. The term hard voting is definite arithmetically by Eq. (4), which requires a statistical mode of classifiers' forecasts.

$$y_i = mode\{c_1, c_2, \ldots, c_k\} \quad (4)$$

However, the hard voting is simple to execute and provides improved performances than baseline classifiers, but, it could not take into thought the probability of slight-based classes.

#### 1) LSTM

LSTM is a kind of ANN planned to procedure sequential data namely text, time sequences, or audio [21]. It is mostly valuable to process data with long-term dependencies, but the outcome at a provided time step is dependent upon data in preceding time steps. LSTM networks have been capable of recollecting this data for lengthy periods by utilizing forget gates, memory cells, output, and input gates. LSTM is generally utilized for tasks like stock price prediction, language translation, and speech detection. Fig. 2 demonstrates the structure of LSTM. The fundamentals of the LSTM network comprise:

Input gate: Control the data flow into the memory cell.

$$inputgate = \sigma(Wf^*[ht-1, xt] + bf) \quad (5)$$

Forget gate: Control the data flow out of the memory cell.

$$forgetgate = \sigma(Wf^*[ht-1, xt] + bf) \quad (6)$$

Output gate: Control the memory cell output to network rest.

$$outputgate = \sigma(Wo^\star[ht-1, xt] + bo) \quad (7)$$

Cell state: Save the data from the memory cell.

$$memorycell = ft^*ct - 1 + it^*tanh(Wc^*[ht-1, xt] + bc) \quad (8)$$

Hidden layer: Resultant of the LSTM unit, utilized to create forecasts or pass data to the following LSTM unit.

$$hiddenlayer = ot^\star tanh(ct) \quad (9)$$



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024

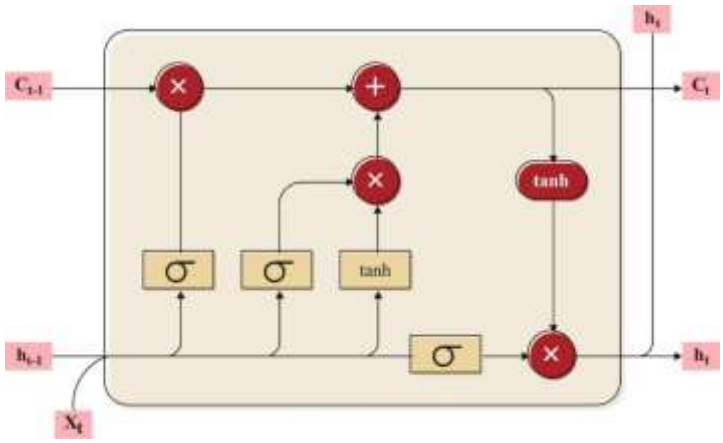


Figure 2: LSTM architecture

**2) ELM**

The ELM is a single implicit layer FFNN model [22]. In comparison to the classical BPNN, ELM needs a random choice of primary weight and threshold, different from the BP model to correct the weight as well as threshold amongst the layers, thereby decreasing the learning time and fundamental difficulty of the model and enhancing the training speed. As part of a combined model, predictive of surface subsidence utilizing ELM could enhance predictive efficacy.

The weight connected between the input and HLs and the threshold value of HL of the ELM are generated randomly during learning and training and could not fine-tuned in succeeding network training. The unique optimum performance can be mainly attained by setting the neuron counts from the HL:

$$\sum_{j=1}^{l} \beta_j \, g\left(\omega_j \alpha_i + \delta_j\right) = d_i \tag{10}$$

$$\omega_j = [\omega_{j1}, \omega_{j2}, \dots, \omega_{jn}]^T \tag{11}$$

$$\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jn}]^T \tag{12}$$

Where $\omega_j$ denotes the weight connected from the input layer to the $j^{th}$ nodes of the HL; $\beta_j$ and $\delta_j$ the output connection weight and threshold of the $j^{th}$ HL nodes; the input and output of the network is $\alpha_i$ and $d_i$, the inner product of $\omega_j$ is $\omega_j \alpha_i$ and the activation function is $\alpha_i$; $g(x)$.

**3) ARNN**

RNN is a variant of traditional NN, extended to address data sequences [23]. While numerous characteristics of NN are retained namely connections and neurons, an RNN can able to repeat a certain function for sequential input through recurrent connection. This efficiently enables to retention of a memory of processed value used along with future input. Assume input series $I = i_1, i_2, i_3, i_T$, for $t$ step, the network repeats the process as follows:

$$\begin{bmatrix} \hat{o}_t \\ h_t \end{bmatrix} = \phi_W(i_t, h_{t-1}) \tag{13}$$

In Eq. (13), $\hat{o}_t$ and $h_t$ are the output and HL at $t$ time, correspondingly. $\phi_W$ is a NN described by weighted network $W$. Assume the $t^{th}$ input $i_t$ and the prior HL $h_{t-1}$ as an input. The architecture of RNN is relatively flexible thus tackling several challenging issues.

The outline of attention mechanism RNN architecture aids the network in preserving memory. This improves performance and enhances memory retention over long sequences. While there are many diverse ways of executing attention, here, an execution of Luong attention is applied.

At $t$ time step, the Luong attention mechanism accounts for the encoder weight $w_t$ over an encoder source series:

$$\sum_{s} w_t\,(s) = 1, and\ \forall s w_t(s) \geq 0 \tag{14}$$

The output value predicted in timestep represents the encoder HL and RNN hidden state $h_t$ as follows:

$$\sum_{s} w_t\,(s) * \hat{h}_s \tag{15}$$

The main difference between attention mechanisms is that $w_t$ value is defined. The Luong attention mechanism exploits the softmax function to calculate sequence score as follows:

$$w_t(s) \leftarrow \frac{exp\left(\beta_t * score\left(h_t, \hat{h}_s\right)\right)}{\sum_s \left(\beta_t * score\left(h_t, \hat{h}_{s'}\right)\right)} \tag{16}$$



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024

In Eq. (16), a scaling control parameter of the attenuation module is $\beta$. The score value for each sequence is defined by the dot product of encoder $\hat{h}_s$ HL and RNN $h_t$ HL transformed through matrix $W_\alpha$:

$$score(h_t, \hat{h}_s) \leftarrow h_t^T W_\alpha \hat{h}_s \quad (17)$$

Here the maximum iteration counter is $T$.

### D. Hyperparameter Tuning

Finally, SCA has been applied for better hyperparameter choice of the DL models. The SCA is a population-based optimizer [24]. The algorithm starts with random population initialization of a promising performance. The exploration and exploitation stages are the 2 different phases in the SCA. In the exploration phase, the algorithm is used to make changes that are more dramatic with a high amount of randomness to localize potential search areas. These changes are more subtle during the exploitation stage, and further randomness has a less beneficial impact.

The two equations upgrade the position of individual performances at the exploration and exploitation since the method exploits the sine and cosine functions during optimization. The position formula for the sine and cosine functions are given below:

$$X_i^{t+1} = X_i^t + r_1 \times sin(r_2) \times |r_3 P_i^t - X_i^t| \quad (18)$$

$$X_i^{t+1} = X_i^t + r_1 \times cos(r_2) \times |r_3 P_i^t - X_i^t| \quad (19)$$

Here is the location of the presented solution at $t^{he\,tth}$ iteration for $the\ i^{th}$ parameter $X_i^t$, $r_1, r_2$ and $r_3$ are random integers, the solution destination for the $i^{th}$ parameter (present optimum individual) is $P_i$, symbol || denotes absolute value.

This equation is combined and utilized as follows:

$$X_i^{t+1} = \begin{cases} X_i^t + r_1 \times sin(r_2) \times |r_3 P_i^t - X_i^t|, & r_4 \geq 0.5 \\ X_i^t + r_1 \times cos(r_2) \times |r_3 P_i^t - X_i^t|, & r_4 < 0.5 \end{cases} \quad (20)$$

with the $r_4$ parameter demonstrating a random integer among 0 and 1, and dictate once the sine or cosine functions are utilized from the search.

The $r_1, r_2, r_3$, and $r_4$ are the four most important parameters in the SCA and dictate the behaviour of the algorithm in specific situations. The $r_1$ parameter directs the movement of next performance that can be towards or away from the presented destination. The $r_2$ parameter is used to dictate the severity of these movements. The $r_4$ parameter is used to present the randomness of the movement, which emphasizes movement if $e_3 < 1$ and deemphasize if $r_3 < 1$. Lastly, the $r_4$ parameter determines whether the sine or cosine functions are applied for the iteration. For the 2D applications, the algorithm is easily extended to cover a high dimensional search range. These behaviors enhance exploitation by concentrating on the spaces among performances. Also, further explorations are needed outside the space within solution, and the SCA obtains this by changing the trigonometric functions or the range. Adjustments to function ranges are adaptively done to balance between exploration and exploitation based on the following expression:

$$r_1 = a0t - \frac{a}{T} \quad (21)$$

Where $t$ denotes the present iteration, $T$ represents the maximal iteration counter and $a$ constant value.

### E. XAI Model: LIME

The developed model chains the XAI technique LIME for an improved explainability and considerate of the black-box model for deceptive content recognition [25]. LIME defines numerous ML methods for regression predictive, utilizing featured value altered of a data sample to convert featured values into the influence of analyst. For the sample, the interpretable method in LIME frequently utilizes linear regression or DTs and is then trained by lesser perturbation (eliminating certain words, hiding image part, and adding random noise) in the technique. The excellence of these methods appeared that growing and applied for resolving optimal share of business victimization databases. Also, there were determined trade-offs among model correctness and interpretability. Mostly, performance is enhanced and greater by using sophisticated models like call trees, boosting, RF, material, and SVM which are “blackbox” methods. The LIME has been applied for some ML black-box methods.

## 4. Experimental Validation

The simulation value of the SCADL-DCDSM methodology takes place on 2 databases [26]. The BuzzFeed dataset holds 273 instances with 2 classes as demonstrated in Table 1. The PolitiFact database comprises 360 instances with 2 classes as displayed in Table 2.

Table 1: Details on BuzzFeed database

| BuzzFeed Database | |
|---|---|
| Classes | No. of Instances |



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024

| | |
|---|---|
| Real News (RN) | 182 |
| Fake News (FN) | 91 |
| Total Instances | 273 |

Table 2: Details on the PolitiFact database

| PolitiFact Database | |
|---|---|
| Classes | No. of Instances |
| Real News | 240 |
| Fake News | 120 |
| Total Instances | 360 |

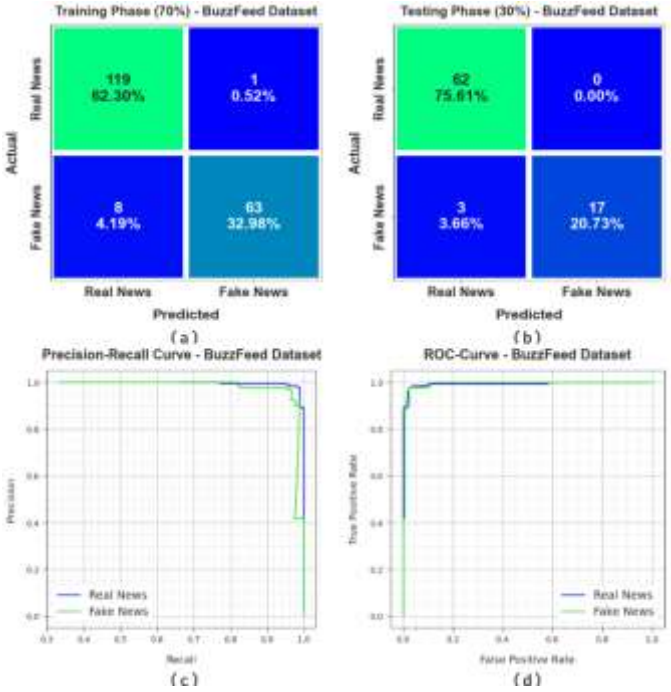


Figure 3: BuzzFeed dataset (a-b) Confusion matrices and (c-d) PR and ROC curves

Fig. 3 represents the classifier performances of the SCADL-DCDSM system on the BuzzFeed database. Figs. 3a-3b illustrates the confusion matrices attained by the SCADL-DCDSM system under 70:30 of TRPH/TSPH. The simulation value implies that the SCADL-DCDSM technique has classified and detected on 2 classes. Besides, Fig. 3c defines the PR result of the SCADL-DCDSM model. The outcome portrayed that the SCADL-DCDSM algorithm has offered the superior solution of PR on 2 classes. Then, Fig. 3d signifies the ROC curve of the SCADL-DCDSM approach. The simulation outcome revealed that the SCADL-DCDSM algorithm has controlled to accomplish performances with maximum outcomes of ROC on 2 classes.

In Table 3 and Fig. 4, the detection performance of the SCADL-DCDSM method on the BuzzFeed database is portrayed. The simulation outcome implied the SCADL-DCDSM system categorized RN and FN. With 70% of TRPH, the SCADL-DCDSM technique obtains an average $accu_y$ of 93.95%, $prec_n$ of 96.07%, $reca_l$ of 93.95%, and $F_{score}$ of 94.84%. Additionally, with 30% of TSPH, the SCADL-DCDSM algorithm gains an average $accu_y$ of 92.50%, $prec_n$ of 97.69%, $reca_l$ of 92.50%, and $F_{score}$ of 94.76%.

Table 3: Detection outcome of SCADL-DCDSM algorithm on BuzzFeed database

| Classes | $Accu_y$ | $Prec_n$ | $Reca_l$ | $F_{Score}$ |
|---|---|---|---|---|
| 70% of TRPH | | | | |
| RN | 99.17 | 93.70 | 99.17 | 96.36 |
| FN | 88.73 | 98.44 | 88.73 | 93.33 |
| Average | 93.95 | 96.07 | 93.95 | 94.84 |



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024

| 30% of TSPH | | | | |
|---|---|---|---|---|
| RN | 100.00 | 95.38 | 100.00 | 97.64 |
| FN | 85.00 | 100.00 | 85.00 | 91.89 |
| Average | 92.50 | 97.69 | 92.50 | 94.76 |

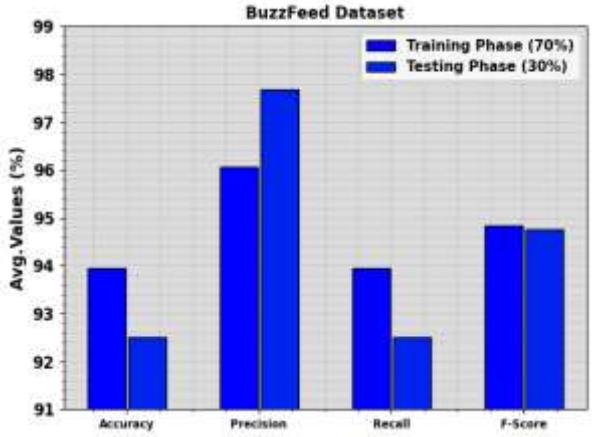


Figure 4: Average of SCADL-DCDSM technique on BuzzFeed dataset

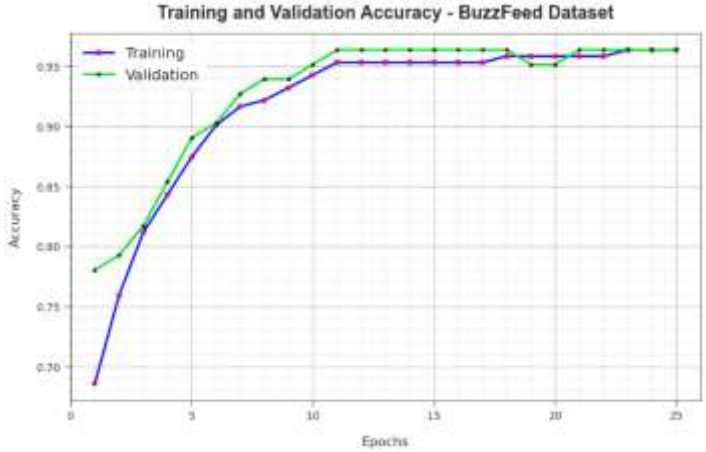


Figure 5: $Accu_y$ curve of SCADL-DCDSM technique on BuzzFeed dataset

The training (TRA) and validation (VL) $accu_y$ curves of the SCADL-DCDSM algorithm on the BuzzFeed dataset displayed in Fig. 5, offer appreciated understandings into the solution of the SCADL-DCDSM algorithm under various epochs. These curves determine the important understandings into the learning method and the model's ability to simplify. In addition, it could be observable that there is a reliability enhancement from TRA and TES $accu_y$ over maximal epochs. It proceedings that the model's ability to learn and distinguish designs from TRA and TES data. The aggregate TES $accu_y$ suggests that the model not only fine-tunes to the TRA data for excels in construction correct predictive on earlier unnoticed data, emphasizing its robust generalized proficiencies.

In Fig. 6, we exemplify a widespread assessment of the TRA and TES loss values for the SCADL-DCDSM approach on the BuzzFeed dataset across different epochs. The TRA loss gradually diminishes as the model better its weighted to minimize classifier errors on together TRA and TES data. These loss curves offer an obvious portrait of how great the model supports with the TRA dataset, highlighting its ability to proficiently hold outlines in both data. It can be worth observing that the SCADL-DCDSM system continually upgrades its parameters to decrease the differences among the prediction and the actual TRA classes.



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024


Figure 6: Loss curve of SCADL-DCDSM methodology on BuzzFeed database

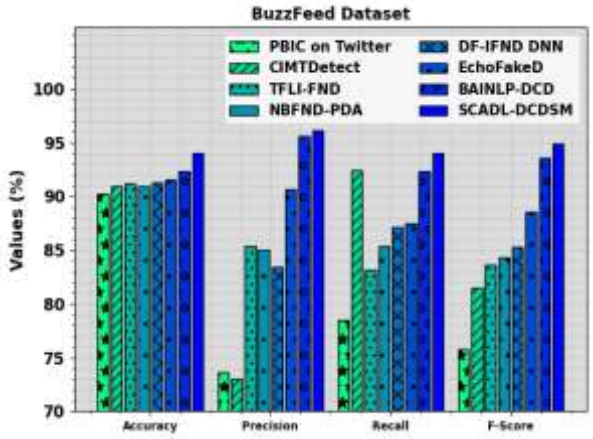


Figure 7: Comparative outcome of SCADL-DCDSM technique under BuzzFeed dataset

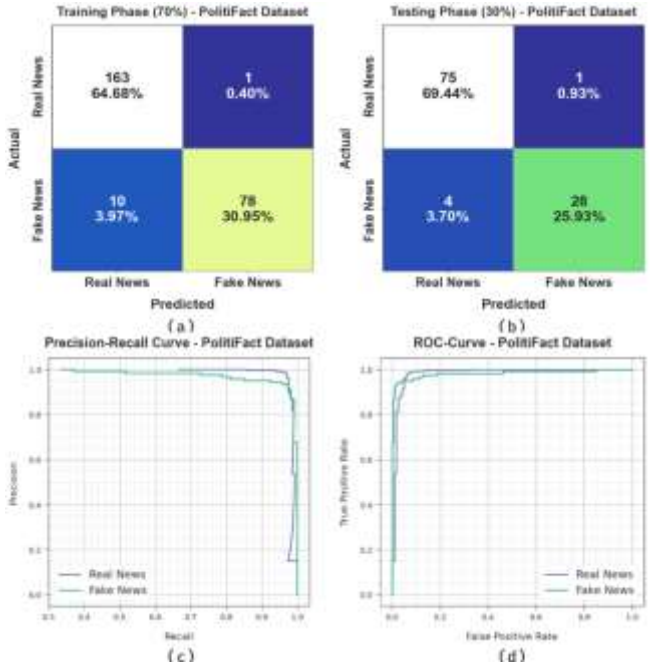


Figure 8: PolitiFact dataset (a-b) Confusion matrices and (c-d) PR and ROC curves

In Fig. 7, the comparison study of the SCADL-DCDSM algorithm reported on the BuzzFeed dataset is provided [27]. The simulation outcome shows that the PBIC, CIMTDetect, and NBFND-PDA algorithms are depicted as



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024

the lowest outcome. Concurrently, the TFLI-FND, DF-IFND DNN, and EchoFakeD approaches have obtained nearby performances. Although the BAINLP-DCD model exhibits considerable outcomes, the SCADL-DCDSM technique accomplishes superior results with a maximum $accu_y$ of 93.95%, $prec_n$ of 96.07%, $reca_l$ of 93.95%, and $F_{score}$ of 94.84%.

Fig. 8 describes the classifier result of the SCADL-DCDSM methodology under the PolitiFact database. Figs. 8a-8b demonstrates the confusion matrices accomplished by the SCADL-DCDSM approach on 70:30 of TRPH/TSPH. The outcome inferred that the SCADL-DCDSM method has categorized and detected 2 classes. Afterward, Fig. 8c demonstrates the PR examination of the SCADL-DCDSM technique. The result denoted that the SCADL-DCDSM model has reached higher results of PR in all classes. However, Fig. 8d represents the ROC outcome of the SCADL-DCDSM technique. The simulation value outperformed that the SCADL-DCDSM algorithm has led to adept performances with greater results of ROC on 2 classes.

In Table 4 and Fig. 9, the detection outcome of the SCADL-DCDSM algorithm on the PolitiFact database is exposed. The outcome implied that the SCADL-DCDSM methodology categorized RN and FN.

Table 4: Detection outcome of SCADL-DCDSM technique on PolitiFact database

| Classes | $\mathbf{Accu_y}$ | $\mathbf{Prec_n}$ | $\mathbf{Reca_l}$ | $\mathbf{F_{Score}}$ |
|---|---|---|---|---|
| TRPH (70%) | | | | |
| RN | 99.39 | 94.22 | 99.39 | 96.74 |
| FN | 88.64 | 98.73 | 88.64 | 93.41 |
| Average | 94.01 | 96.48 | 94.01 | 95.07 |
| TSPH (30%) | | | | |
| RN | 98.68 | 94.94 | 98.68 | 96.77 |
| FN | 87.50 | 96.55 | 87.50 | 91.80 |
| Average | 93.09 | 95.74 | 93.09 | 94.29 |

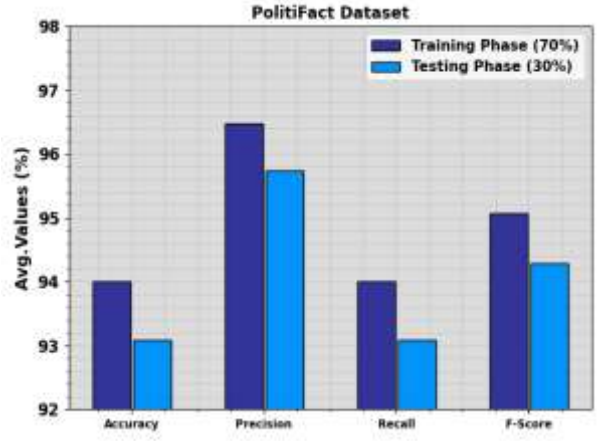


Figure 9: Average of SCADL-DCDSM technique on PolitiFact dataset

With 70% of TRPH, the SCADL-DCDSM methodology reaches an average $accu_y$ of 94.01%, $prec_n$ of 96.48%, $reca_l$ of 94.01%, and $F_{score}$ of 95.07%. Furthermore, with 30% of TSPH, the SCADL-DCDSM system accomplishes an average $accu_y$ of 93.09%, $prec_n$ of 95.74%, $reca_l$ of 93.09%, and $F_{score}$ of 94.29%.



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024

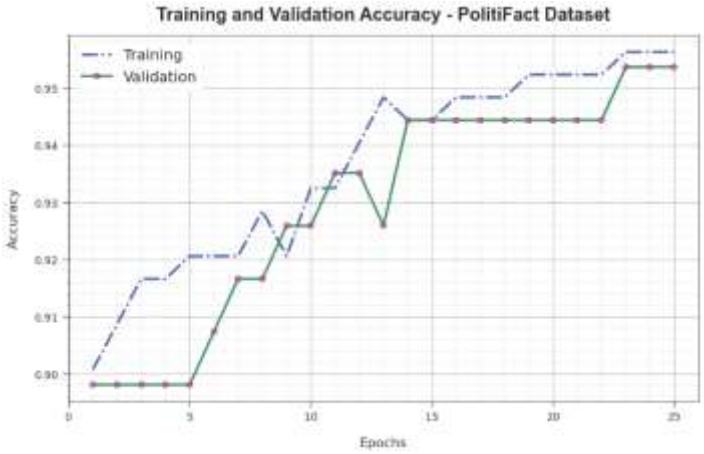


Figure 10: $Accu_y$ curve of SCADL-DCDSM methodology on PolitiFact database

The TRA and VL $accu_y$ curves of the SCADL-DCDSM algorithm on the PolitiFact database exposed in Fig. 10, suggest valued perceptions into the performance of the SCADL-DCDSM algorithm under various epochs. In addition, it could be observable that there is a dependability upgrading from TRA and TES $accu_y$ over higher epochs. It is noticeable the model's capability to acquire and distinguish designs from the TRA and TES data. The maximum TES $accu_y$ presents that the model not only fine-tunes to the TRA data for excels in creating accurate predictive on earlier unnoticed data, emphasizing its robust generalized capabilities.

In Fig. 11, we suggest a wide-ranging assessment of the TRA and TES loss values for the SCADL-DCDSM technique on the PolitiFact dataset under many epochs. The TRA loss gradually drops as the model develops its weights to minimize classifier errors on TRA and TES data. These loss curves offer an obvious representation of how great the model supports with the TRA data, emphasizing its ability to proficiently hold designs in both databases. It can be worth observing that the SCADL-DCDSM methodology continually enhances its parameters to decrease the differences among the predictive and the actual TRA class.

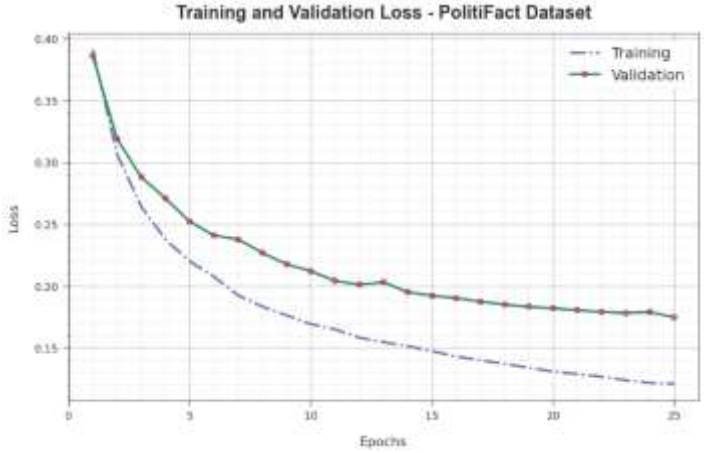


Figure 11: Loss curve of SCADL-DCDSM method on PolitiFact database

In Fig. 12, the comparative outcome of the SCADL-DCDSM system is reported on the PolitiFact database. The experimental value implied that the PBIC and CIMTDetect algorithms have outperformed the least solution. Also, the TFLI-FND, DF-IFND DNN, and EchoFakeD methodologies have attained nearby performances. But the BAINLP-DCD system exhibits considerable outcomes, the SCADL-DCDSM algorithm realizes superior performances with maximal $accu_y$ of 94.01%, $prec_n$ of 96.48%, $reca_l$ of 94.01%, and $F_{score}$ of 95.07%.



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024

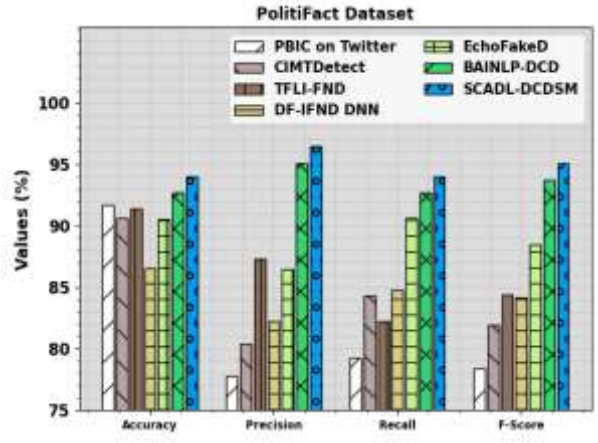


Figure 12: Comparative outcome of SCADL-DCDSM technique under PolitiFact dataset

These results ensured that the SCADL-DCDSM algorithm accomplishes improved solution on deceptive content detection.

## 6. Conclusion

In this study, we have presented a novel SCADL-DCDSM methodology. The SCADL-DCDSM technique incorporates the ensemble learning process with hyperparameter tuning strategy for classifying the sentiments. Primarily, the SCADL-DCDSM technique pre-processes the input data to change the input data into the beneficial format. Also, the SCADL-DCDSM technique follows the BERT model for the word embedding process. Moreover, the SCADL-DCDSM technique involves an ensemble of three models for sentiment classification such as LSTM, ELM, and ARNN. Finally, SCA can be executed for better hyperparameter choice of the DL models. The SCADL-DCDSM algorithm integrates the XAI system LIME and is employed for a comprehensive explainability and understanding of the black-box process, enhancing correct deceptive content recognition. The simulation value of the SCADL-DCDSM algorithm has been examined on a benchmark database. The experimental values illustrated that the SCADL-DCDSM technique achieves a better solution than other approaches with respect to different measures.

**Funding:** "This research received no external funding"

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

## References

[1] Sahoo, S.R. and Gupta, B.B., 2021. Multiple features based approach for automatic fake news detection on social networks using deep learning. Applied Soft Computing, 100, p.106983.
[2] Okunoye, O.B. and Ibor, A.E., 2022. Hybrid fake news detection technique with genetic search and deep learning. Computers and Electrical Engineering, 103, p.108344.
[3] Kaliyar, R.K., Goswami, A. and Narang, P., 2021. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia tools and applications, 80(8), pp.11765-11788.
[4] Chung, W., Zhang, Y. and Pan, J., 2023. A theory-based deep-learning approach to detecting disinformation in financial social media. Information Systems Frontiers, 25(2), pp.473-492.
[5] Tashtoush, Y., Alrababah, B., Darwish, O., Maabreh, M. and Alsaedi, N., 2022. A deep learning framework for detection of COVID-19 fake news on social media platforms. Data, 7(5), p.65.
[6] Sarkar, S., Tudu, N. and Das, D., 2023. HIJLI-JU-CLEF at MULTI-Fake-DetectiVE: Multimodal Fake News Detection Using Deep Learning Approach. In Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), CEUR. org, Parma, Italy.
[7] Sarnovský, M., Maslej-Krešňáková, V. and Ivancová, K., 2022. Fake news detection related to the covid-19 in slovak language using deep learning methods. Acta Polytechnica Hungarica, 19(2), pp.43-57.
[8] Ghanem, R. and Erbay, H., 2023. Spam detection on social networks using deep contextualized word representation. Multimedia Tools and Applications, 82(3), pp.3697-3712.
[9] Birunda, S.S., Nagaraj, P., Narayanan, S.K., Sudar, K.M., Muneeswaran, V. and Ramana, R., 2022, January. Fake Image Detection in Twitter using Flood Fill Algorithm and Deep Neural Networks. In 2022 12th



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024

International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 285-290). IEEE.
[10] Maathuis, C. and Godschalk, R., 2023, February. Social Media Manipulation Deep Learning based Disinformation Detection. In International Conference on Cyber Warfare and Security (Vol. 18, No. 1, pp. 237-245).
[11] Hamed, S.K., Ab Aziz, M.J. and Yaakub, M.R., 2023. DISINFORMATION DETECTION ABOUT ISLAMIC ISSUES ON SOCIAL MEDIA USING DEEP LEARNING TECHNIQUES. Malaysian Journal of Computer Science, 36(3), pp.242-270.
[12] Alouffi, B., Alharbi, A., Sahal, R. and Saleh, H., 2021. An optimized hybrid deep learning model to detect COVID-19 misleading information. Computational Intelligence and Neuroscience, 2021.
[13] Fouad, K.M., Sabbeh, S.F. and Medhat, W., 2022. Arabic Fake News Detection Using Deep Learning. Computers, Materials & Continua, 71(2).
[14] Sadiq, S., Aljrees, T. and Ullah, S., 2023. Deepfake Detection on Social Media: Leveraging Deep Learning and FastText Embeddings for Identifying Machine-Generated Tweets. IEEE Access.
[15] Fan, H., Du, W., Dahou, A., Ewees, A.A., Yousri, D., Elaziz, M.A., Elsheikh, A.H., Abualigah, L. and Al-qaness, M.A., 2021. Social media toxicity classification using deep learning: real-world application UK Brexit. Electronics, 10(11), p.1332.
[16] Kaliyar, R.K., Goswami, A., Narang, P. and Chamola, V., 2022. Understanding the Use and Abuse of Social Media: Generalized Fake News Detection With a Multichannel Deep Neural Network. IEEE Transactions on Computational Social Systems.
[17] Prakash, O. and Kumar, R., 2023, January. Fake News Detection in Social Networks Using Attention Mechanism. In Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 2 (pp. 453-462). Singapore: Springer Nature Singapore.
[18] Kumar, A. and Sachdeva, N., 2022. A Bi-GRU with attention and CapsNet hybrid model for cyberbullying detection on social media. World Wide Web, 25(4), pp.1537-1550.
[19] Atandoh, P., Zhang, F., Adu-Gyamfi, D., Atandoh, P.H. and Nuhoho, R.E., 2023. Integrated deep learning paradigm for document-based sentiment analysis. Journal of King Saud University-Computer and Information Sciences, 35(7), p.101578.
[20] Mohammed, A. and Kora, R., 2022. An effective ensemble deep learning framework for text classification. Journal of King Saud University-Computer and Information Sciences, 34(10), pp.8825-8837.
[21] Gülmez, B., 2023. Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm. Expert Systems with Applications, 227, p.120346.
[22] Wang, Y., Dai, F., Jia, R., Wang, R., Sharifi, H. and Wang, Z., 2023. A novel combined intelligent algorithm prediction model for the tunnel surface settlement. Scientific Reports, 13(1), p.9845.
[23] Predić, B., Jovanovic, L., Simic, V., Bacanin, N., Zivkovic, M., Spalevic, P., Budimirovic, N. and Dobrojevic, M., 2023. Cloud-load forecasting via decomposition-aided attention recurrent neural network tuned by modified particle swarm optimization. Complex & Intelligent Systems, pp.1-21.
[24] Bacanin, N., Jovanovic, L., Zivkovic, M., Kandasamy, V., Antonijevic, M., Deveci, M. and Strumberger, I., 2023. Multivariate energy forecasting via metaheuristic tuned long-short term memory and gated recurrent unit neural networks. Information Sciences, 642, p.119122.
[25] Patil, S., Varadarajan, V., Mazhar, S.M., Sahibzada, A., Ahmed, N., Sinha, O., Kumar, S., Shaw, K. and Kotecha, K., 2022. Explainable artificial intelligence for intrusion detection system. Electronics, 11(19), p.3079.
[26] https://www.kaggle.com/datasets/mdepak/fakenewsnet
[27] Albraikan, A.A., Maray, M., Alotaibi, F.A., Alnfiai, M.M., Kumar, A. and Sayed, A., 2023. Bio-Inspired Artificial Intelligence with Natural Language Processing Based on Deceptive Content Detection in Social Networking. Biomimetics, 8(6), p.449.



Doi: https://doi.org/10.54216/FPA.140212
Received: July 18, 2023 Revised: November 18, 2023 Accepted: January 16, 2024