Integrating Metaheuristics and Two-Tiered Classification for Enhanced Fake News Detection with Feature Optimization

Poonam Narang[1,*], Ajay Vikram Singh[2] and Himanshu Monga[3]

[1,2]AIIT, Amity University, Noida, India
[3]Govt. Hydro Engineering College, Bandla, India

Abstract

INTRODUCTION: The challenge of distributing false information continues despite the significant impact of social media on opinions. The suggested framework, which is a metaheuristic method, is presented in this research to detect bogus news. Employing a hybrid metaheuristic RDAVA methodology coupled with Bi-LSTM, the method leverages African Vulture Optimizer and Red Deer Optimizer.

OBJECTIVES: The objective of this study is to assess the effectiveness of the suggested model in identifying false material on social media by employing social network analysis tools to combat disinformation.

METHODS: Employing the data sets from BuzzFeed, FakeNewsNet, and ISOT, the suggested model is implemented on the MATLAB Platform and acquires high accuracy rates of 97% on FakeNewsNet and 98% on BuzzFeed and ISOT. A comparative study with current models demonstrates its superiority.

RESULTS: Outperforming previous models with 98% and 97% accuracy on BuzzFeed/ISOT and FakeNewsNet, respectively, the suggested model shows remarkable performance.

CONCLUSION: The proposed strategy shows promise in addressing the problem of false information on social media in the modern day by effectively countering fake news. Its incorporation of social network analysis methods and metaheuristic methodologies makes it a powerful instrument for identifying false news.

1. Introduction

The proliferation of false news on social media will severely impact every domain. The social media platform should facilitate the identification scheme that distinguishes fake news from original data. As we know, social media relies on the Internet, but before the Internet was even a thing, there were hoaxes and fake news [1] [2]. Fake news on the Internet is generally defined as "fictitious stories that have been purposefully created to deceive readers." To promote reading or as a kind of psychological warfare, news organizations and social media sites disseminate false information. However, distinguishing fake and genuine data from social media platforms is highly complicated due to the exponential proliferation of information [3] [4] [5]. As a result, social media is now contaminated with fake data; therefore, a requirement is to develop a highly efficient model to detect fake news from social media sites.

The detection model faces severe complications due to a lack of tagged data identifying fake news [6]. Fake news can be

*Corresponding author. Email: hipoonam@gmail.com

broadly categorized as clickbait, misinformation, disinformation, hoaxes, satire, parody, deceptive news, rumours, and similar data [7] [8]. Fake news on social media is circulated several times, leading to substantial detrimental effects on society due to misleading information [9] [10]. Social media-based fake news identification is becoming a hot topic where the quick circulation of data may contain misleading material [11] [12]. The spread of false information in today's media environment has become a prevalent issue with far-reaching effects on society. For instance, a lot of misinformation spread extensively on social media platforms during the 2016 U.S. presidential election [48], affecting public views and influencing voting behavior. These events made clear how fake news may influence political discourse and jeopardize the democratic process. Likewise, in the field of public health, the dissemination of false information about COVID-19 therapies and vaccinations [49] has presented significant challenges for medical experts and decision-makers. Online conspiracies and false information have weakened attempts to contain the pandemic and increased vaccine mistrust.

Thus, identifying false news is becoming inevitable and getting extreme attention from every field [13]. However, the existing models cannot perform well in detecting fake news due to the complexity of identifying similar data [14] [15]. For better identification, the auxiliary information related to individual social media contribution because the fake news primarily misguided the individual into false information. The detection model performs only with the existing data but does not work well with the newly arrived data. Second, supplementary data may be complex due to significant, chaotic, sparse, and noisy data. Therefore, detecting fake news is highly complicated on social media datasets.

Therefore, the contribution of this work is to analyze the recent research works n fake news detection and to define the problem statement of the same. Then the significant features from the pre-processed data for efficient data categorization (fake/genuine). After that, the most relevant features among the extracted features with a new hybrid optimization model to reduce the computational complexity of the model. Finally, a new two-level deep-learning classifier model for accurate data classification can be developed—the hyper-parameters of the deep-learning model with a new hybrid optimization model to process fine-tuning—moreover, validation for proving the performance efficiency of the projected model.

1.1 Research Objective

The primary objective of this study is to develop and apply a reliable model for identifying false information on social media websites. The study intends to investigate how social media affects both individual and group viewpoints, with a particular emphasis on the spread of false and true information. The main research objectives of the proposed approach are:

1. Examine how individual and collective viewpoints are susceptible to the impact of social media, with a focus on the novel aspects of this influence.
2. Highlighting the newness of these concerns, examine the particular difficulties brought about by the ease with which false information can proliferate on social media.
3. Create a novel metaheuristic approach for the inventive identification of bogus news by combining African Vulture Optimisation and Red Deer Optimisation.
4. Utilise a unique hybrid RDAVA approach and Bi-LSTM classifier for feature selection and two-level classification to improve accuracy.

2. Literature Review

The literature review section mainly provides different authors' approaches that helped classify or predict fake news. The existing techniques and their performance on other datasets are highly considerable criteria for developing novel courses.

The semantic structure of text data can only be understood with Natural language processing (NLP)[45][46]activities such as text classification, sentiment analysis, POS tagging, and machine translation. Through the analysis of syntactic structures and contextual information, these tasks aid in the differentiation between various word senses. For example, classification helps to discern between fake and real news by classifying documents according to their content. Sentiment analysis helps to discern the sentiments conveyed in text, and POS tagging helps to distinguish between word senses by giving words grammatical labels and improving the capacity to distinguish between real and false news.

NLP tasks including text categorization, sentiment analysis, and named entity recognition have been transformed by recent advances in deep learning, best demonstrated by transformer-based designs like BERT [42][43], GPT[44], and their variations. These architectures find complex linguistic patterns and semantic subtleties in text data by utilising massive pre-trained language models and advanced attention processes. Furthermore, models of neural networks with attention mechanisms added to them have shown to be remarkably effective in modelling long-range relationships and encoding contextual information, which has resulted in more precise predictions and a better comprehension of textual semantics.

Drawing insights from studies [38][39][40][41] on distributed database optimization, proactive detection of online antisocial behavior, and crop yield prediction using deep learning and feature selection techniques, our hybrid metaheuristic approach for enhanced fake news detection with feature optimization is informed by these sources.

Therefore, we have collected research articles from 2019 to 2023. Consequently, we have reviewed 20 recent research articles based on datasets, performance, and techniques. Several techniques were developed to detect Fake news based on several datasets. The literature summary is tabulated in

Table 1, which helps to understand all the techniques in the existing approaches and datasets.

3. Proposed Methodology

The Online Social Media platform is a prevalent and powerful tool to transfer data to all users. However, the complex data environment of social media may provide the chance of spreading false news in terms of deliberate, false, or misleading information; therefore, a highly enhanced model is required to identify fake news. Accordingly, a novel fake news detection model has been developed to detect fake news in social media networks. The proposed model's architecture is shown in Fig.1. Initially, the dataset description is discussed in the following section.

Table 1. Summary of Literature

Ref	Year	Technique	Dataset	Result
[16]	2022	FND-NS	NELA-GT-19 and Fakeddit	74% of accuracy
[17]	2022	CNN based CredNN	FakeNewsNet and PolitiFact	93.7% % of Accuracy
[18]	2022	CB-Fake	Politifact and Gossipcop	93% of accuracy
[19]	2021	Machine Learning-based PU-LP	FakeNewsNet and Fake.BR	F1 measure based on Avg ± Std is 8.8 ± 4.6 for BoW and 6.2 ± 3.2 for D2V
[20]	2022	FNID-LIAR	FakeNewsNet and LIAR datasets	85.58% of accuracy
[21]	2021	DeepFakE	BuzzFeed	85.86% of accuracy
[22]	2021	Linguistic Feature-Based Learning Model	FakeNewsNet and BuzzFeed	86% of accuracy
[23]	2021	CNN-RNN	ISO and FA-KES	60% of Accuracy for Fa-KES
[24]	2019	CNN + bidirectional LSTM	Twitter and PolitiFact	88.78% of accuracy
[25]	2021	MVAN	Twitter15 and Twitter16	90%
[26]	2021	EchoFakeD	BuzzFeed and PolitiFact	92.30%
[27]	2021	AMFB	Twitter and Weibo	Accuracy for Twitter is 88%, and for Weibo, 89%
[28]	2021	LSTM+VGG16	IFND dataset	82%
[29]	2021	DAGA-NN	Twitter and Weibo	Accuracy for Twitter is 90% and for Weibo is 94%
[30]	2020	CNN-LSTM	Fake News Challenges (FNC)	97.8% accuracy
[31]	2020	LSTM-CNN	Twitter, and real-time online users.	92.89% of accuracy
[32]	2021	Light CNN, RestNet-101, Xception	Deepfake dataset FaceForensics++	90% of accuracy
[33]	2020	FakeBERT	Real-time dataset based on U.S. General Presidential Election-2016	98% of accuracy

3.1 Dataset description

The implementation of the proposed model is executed with three different datasets: Buzzfeed Dataset, ISOT dataset, and FakeNewsNet.

Buzzfeed dataset: The Buzzfeed dataset is based on a sample of Facebook news related to the U.S. election held in 2016. Moreover, five BuzzFeed journalists examined four days of news in this dataset. These datasets are CSV files, where every file has 91 observations and 12 features or variables. This dataset incorporates exploratory data analysis part and classification part. In addition, the classification can be done in terms of Fake or Real news [35].

ISOT Dataset: The ISOT dataset is similar to the BuzzFeed dataset, which is collected from news websites related to PolitiFact. The datasets have two different files such as True.CSV and Fake. CSV. Each dataset has 12600 from reuter.com. Moreover, the dataset is mainly from 2016 to 2017 to the data collected Kaggle dataset [36].

FakeNewsNet Dataset: Similar to the ISOT dataset, the FakeNewsNet Dataset is based on PolitiFact. In addition, the dataset is also related to GossipCop. Moreover, this dataset has id, URL, title, and tweet_ids attributes [37].

All these datasets need to be analyzed to predict and identify fake news as it contains a lot of junk data. Therefore, the dataset is processed with a significant data pre-processing.
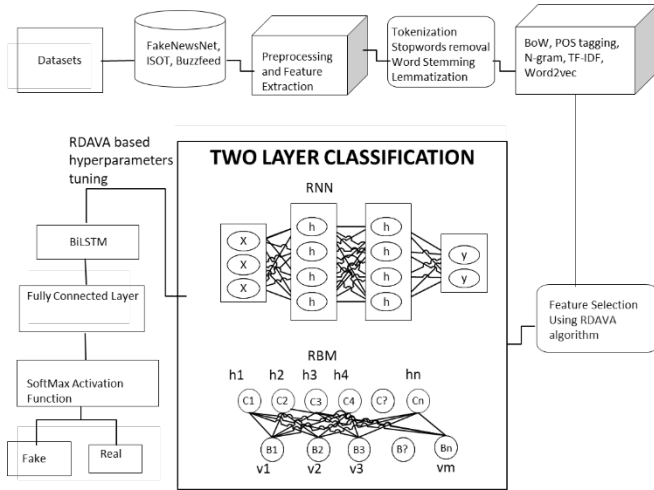
Fig.1 Overall architecture of the proposed model

3.2 Pre-processing

The pre-processing helps to reduce the dimensionality of the collected raw data where various steps like sentence segmentation, tokenization, stopwords removal, and word stemming, as discussed below in detail: Initially, the input sentences are separated by utilizing certain operators such as colon (:), dot (.), semicolon (;), etc. In addition, the white space and tab key can also be utilized to divide sentences.

By using all these keys, the sentence can be transferred into tokens. Then the repeated words and stop words are identified for removing them to make more clarity in classification. For instance, the words such as on, in, at, thus, and too, a, and an are some stop words of the sentence. These words do not have any specific meaning, so the removal of these will not impact performance. Then the stemming process may help to reform the word to its basic form. After all this process, the feature extraction is processed, and the detailed description of the feature extraction process is discussed in the following section.

3.3 Feature Extraction

The computation complexity is always higher in the learning process with high dimensional data. The data may be associated with a large volume of phrases and words that make learning complex. On the other hand, the learning performance may be digested because of redundancy and irrelevant data. Therefore, an appropriate feature selection algorithm is required for feature reduction to reduce the size and overcome ample dimensional feature space.

Initially, the Bag of Words (BoW) converts textual data into vectors of word frequencies or occurrences, providing a numerical representation of documents. This representation is then utilized in various NLP tasks such as text classification, sentiment analysis, and document clustering by applying machine learning algorithms that operate on these vectorized representations to make predictions or extract insights from the text data. Then the POS tags for

statistical NLP tasks since they can be used to discriminate between different word senses, which is highly useful for realizing texts and deriving semantic information from given texts. Then the incorporation of the N-gram is to illustrate the context of the document and provide features to classify the document. Moreover, it is utilized to formulate several sets of n-gram frequency profiles from the training data to represent fake and truthful news articles. The I-TF-IDF has been used with the RDAVA algorithm for further processing, i.e., clearly stated as follows.

I-TF-IDF using RDAVA

Improved Term Frequency-Inverse Document Frequency, i.e., I-TF-IDF, is based on evaluating the fitness function. Therefore, the Mean Absolute Difference (MAD) is considered the Fitness Function, i.e., assessed using the RDAVA algorithm. The I-TF-IDF helps to select more significant features. The weight of the TF-IDF algorithm is supposed to update the essential features using the following equation.

$$MAD_{xi} = \frac{1}{a_i} \sum_{j=1}^{t} |x_{ij} - \hat{x}_i| \qquad (1)$$

Where,

$$\hat{x}_i = \left(\frac{1}{a}\right) \sum_{j=1}^{t} x_{ij} \qquad (2)$$

Where x_{ij}, Represents the feature value in which "i" is a textual document and j is a feature. The selected features are described as a_i, t is the unique features, and $\hat{x}_i$ Represents the mean of vector i. The Word2Vec generates vectors of the words that are distributed numerical representations of word features. These word features encapsulate words that illustrate the context of each word present in the sentence. Word embedding finally helps establish the collection of a word with another similar meaning word through the created vectors. Later, the feature selection with the RDAVA algorithm is obtained and discussed in detail below.

RDAVA-based Feature Selection

The pre-processed data based on several word-level processing, including TF-IDF, are considered. Moreover, the TF-IDF-based feature extraction approach is further enhanced by incorporating a novel RDAVA algorithm. The RDAVA algorithm emerged from the idea of two different algorithms. The RDAVA algorithm emerged from the idea of two different algorithms- Red deer optimization (RDO) and African Vulture Optimization Algorithm (AVOA).

RDO [47] exhibits an equilibrium between exploration and exploitation when pursuing the optimal solutions; it was inspired by the red deer's foraging behavior. RDO is appropriate for a variety of optimisation problems because of its behavior, which enables it to travel solution spaces effectively. RDO exhibits competitive convergence speed, solution quality, and scalability when measured against other metaheuristic algorithms such as Genetic Algorithms (GA) or Particle Swarm Optimisation (PSO). For instance, robust convergence to near-optimal solutions is regularly demonstrated by RDO in optimisation problems with high-dimensional search spaces. Its adaptability in modifying its search approach according to the specifics of the problem also adds to its efficacy in a variety of optimisation sectors.

The following shows detailed information about the AVOA algorithm.

Background model of the AVOA

AVOA algorithm is inspired by the behavior of African vultures that execute specific strategies for hunting their prey [33]. The following steps describe the process of AVOA.

Rule 1: The population of the African vultures is denoted as N, and D represents the position space or dimension of each vulture. In addition, D alas dimension of the problem may vary regarding different issues. Likewise, T as maximum iteration, i.e., helps to solve the complex problem. Therefore, "i" denotes each vulture's position and is determined based on condition $1 \leq i \leq N$ in different iterations as $t(1 \leq t \leq T)$. This can be defined using the equation. (3)

$$X_i^t = x_{i1}^t, \dots \dots, x_{id}^t, \dots \dots, x_{iD}^t \quad (3)$$

Rule 2: The entire population of the vultures is divided into three categories, where the first is determined as the best feasible solution from the whole group. The second is the second-best solution, and the rest are denoted as the third group.

Rule 3: The vultures work together to hunt their prey by taking the different leads in the population.

Rule 4: The best and worst vultures in the population are represented by analyzing the fitness values of the hungriest and weakest. Conversely, the strongest vulture plays a leading role and best as well. In the populations of African vultures, the vultures in the population always depend on the strongest vultures, whereas they stay far from the weakest vultures. By analyzing the steps mentioned above of African Vulture, the AVOA has developed five stages of execution strategy.

First Phase: This phase has executed rule 2 to form grouping based on quality, which must be performed after initialization. The first and second are grouped as the first and second best solutions. And the remaining vultures are gathered into the third group. Equation (4) is designed to determine the vulture that should be moved into the current iteration.

$$R_i^t = \begin{array}{l} BV_1^t, p_i^t = L_1 \\ BV_2^t, p_i^t = L_2 \end{array} \quad (4)$$

In equation (2), the first best vultures are mathematically represented as $BV_1^t = b_{11}^t, \dots, b_{1d}^t, \dots, b_{1D}^t$. Likewise, the second-best is denoted as $BV_2^t = b_{21}^t, \dots, b_{2d}^t, \dots, b_{2D}^t$. Then the two random numbers denoted as L1 and L2, are placed between 0 and 1. Based on the Roulette Wheel

Strategy (RWS), the sum of two numbers is obtained as 1. The mathematical formulation of RWS is shown below.

$$P_i^t = \frac{f_i^t}{\sum_{i=1}^{m} f_i^t} \quad (5)$$

Where f_i^t Indicates the fitness value, i.e., determined from the i^{st} (α) group and 2nd (β) group vultures, and m denotes the total number of I^{st} (α) group and 2nd (β) group vultures. The γ in the group represents the weakest, the remaining vultures in the population. After all the identification, the target vulture is identified by applying the significant parameters.

Second Phase: The Hunger of Vultures
The second phase is based on a hungry vulture, which does not have significant strength to find food from a farther location. Therefore, the hungry vulture's behavior is aggressive and stays together with a vulture. Thus exploration and exploitation can be developed based on the degree of hunger. Equation (6) helps to determine the degree of hunger of the ith at the tth iteration.

$$F_i^t = (2 \times rand_{i1}^t + 1) \times z^t \times \left(1 - \frac{t}{T}\right) + g^t \quad (6)$$

Where $rand_{i1}^t$, is a random number in the range of [0,1], z is a random number in the range of [−1,1], and g is calculated by Eq (5).

$$g^t = h^t \times \left(sin^k\left(\frac{\pi}{2} \times \frac{t}{T}\right) + cos\left(\frac{\pi}{2} \times \frac{t}{T}\right) - 1\right) \quad (7)$$

Where h represents the random number between the range of −2,2 and k as the probability of the exploitation stage, a larger k denotes the final optimization stage, i.e., to determine the higher possibility of entering the exploration stage. Conversely, the smaller k determines to enter into the exploitation phase. According to the formula, as long as the number of iterations is increasing, f_i^t will get decreased. In this process, f_i^t is exceeded above 1, the exploration process is executed, and looking for a new location to find food is. It is converse when the f_i^t get below 1 that executes exploitation where food will be identified from the current location.

Phase 3: Exploration Stage
The AVOA has provided two exploration phases, i.e., incorporated with the parameter P1 to make a significant decision at the current time. The P1 has initialized with a specific range between 0 and 1.

$$X_i^{t+1} = \begin{cases} R_i^t - D_i^t \times F_i^t, p_1 \geq rand_{p1}^t \\ (R_i^t - F_i^t + rand_{i2}^t \times \\ \left((ub - lb) \times (rand_{i3}^t + lb), p1 < rand_{p1}^t\right) \end{cases} \quad (8)$$

Here X_i^{t+1} denotes i^{th} vulture's position at the $t+1_{th}$ iteration $rand_{p1}^t$, $rand_{i2}^t$, and $rand_{i3}^t$ are three different random numbers set between the range of 0 and 1. Equation (4) and equation (4) are respectively utilized for obtaining R_i^t and F^t. The upper and lower bounds are represented as ub and lb of

the problem solution. D_i^t is determined by equation (9), which helps to determine the distance between the vulture and the new optimal vulture.

$$D_i^t = |C \times R_i^t - X_i^t| \qquad (9)$$

Where X_i^t represents the ith vulture position in iteration t, and C is uniformly distributed based on a random number in the range [0,2].

Phase 4: Exploitation Stage
The contribution of this phase 4 is highly inevitable in maintaining the balance between exploration and exploitation. Here, $|F_i^t|$ is obtained between 0.5 and 1, leading to the exploitation phase, especially medium-term based. The medium-term based exploitation phase is considered the parameter, i.e., $p2$ with the range of 0 and 1. In the medium-term exploitation stage, a parameter $p2$ utilized to execute rotating flight or food rivalry by considering the range between 0 and 1. These $rand_{p2}^t$ random numbers are generated before executing the vultures act. If the $rand_{p2}^t \geq p2$, the food competition process will be executed. Conversely, the rotating flight will be performed if the $rand_{p2}^t \leq p2$.

a) Food Competition

The strongest vultures have the $|F_i^t|$ the range between 0.5 and 1, and they avoid the weakest vulture from sharing the food, but the weakest one competes with the strongest to get food. This process is mathematically derived using equation (10)

$$X_i^{t+1} = D_i^t \times (F_i^t + rand_{i4}^t) - d_i^t \qquad (10)$$

Where the equation (9) helps to determine D_i^t, the equation (4) to determine F_i^t, $rand_i^t$ is a uniformly distributed random number between the range of 0 and 1. At last the d_i^t is determined using equation (11)

$$d_i^t = R_i^t - X_i^t \qquad (11)$$

b) Rotating Flight
As far as the vulture is full and energetic, it processes food competition and hovers at high altitudes. The vultures are executing spiralling in a rotating fight where they update their position based on equation (12).

$$X_i^{t+1} = R_i^t - (S_{i1}^t + S_{i2}^t) \qquad (12)$$

Where S_{i1}^t and S_{i2}^t are mathematically denoted using the following (13) and (14) equations

$$S_{i1}^t = R_i^t \times \left(\frac{rand_5^t \times X_i^t}{2\pi}\right) \times \cos(X_i^t) \qquad (13)$$

$$S_{i2}^t = R_i^t \times \left(\frac{rand_6^t \times X_i^t}{2\pi}\right) \times \sin(X_i^t) \qquad (14)$$

Phase 5: Exploitation stage (later)
As long as the vultures $|F_i^t|$ is come below 0.5, almost the entire population of the vulture is becoming full, but the best two vultures in the group are becoming weak and starving due to long-term practice. In this scenario, the vultures are gathered in the exact location and attack their prey. In this process, $p3$ is utilized in between the range of 0 and 1. Therefore, the vulture's behavior can be either aggregation or attacking. Thus, the vultures, while reaching into exploitation with $rand_{p3}^t$, as a random number with a

specific range [0,1], these numbers will be generated before the vulture act. Moreover, if $rand_{p3}^t \leq p3$, the attacking behavior will be executed. Conversely, $rand_{p3}^t \geq p3$ the aggregation behavior will be taken place.

a) Aggregation Behavior

The vultures can consume an abundance of food while they are at the late stage. Many vultures are gathered at the food location and execute the competition behavior. Equation (13) is used to update the vulture's position.

$$X_i^{t+1} = \frac{A_{i1}^t + A_{i2}^t}{2} \qquad (15)$$

Where the equation (14) and (15) are utilized to determine A_{i1}^t and A_{i2}^t respectively.

$$A_{i1}^t = \frac{BV_1^t \times X_i^t}{BV_1^t \times (X_i^t)^2} \times F_t^t \qquad (16)$$

$$A_{i2}^t = BV_2^t - \frac{BV_2^t \times X_i^t}{BV_2^t \times (X_i^t)^2} \times F_t^t \qquad (17)$$

b) Attack Behavior
Similar to aggregation behavior, the attack behavior will be executed until it reaches the last stage, but the vultures are moving to the food location to get the abandoned food, i.e., determined by equation (16).

$$X_i^{t+1} = R_i^t - |d_i^t| \times F_i^t \times Levy(dim) \qquad (18)$$

Where, d_i^t is based on equation (11), dim denotes the dimension of the problem. $Levy(.)$ denotes the levy flight, i.e., derived using the equation (19)

$$Levy(dim) = 0.01 \times \frac{r_1 \times \sigma}{|r2|^{\frac{1}{\beta}}} \qquad (19)$$

Where r1 and r2 are uniformly distributed random numbers between the range of 0 and 1, δ is a constant with a value of 1.5. The equation (18) to determine the σ.

$$\sigma = \left(\frac{\Gamma(1+\beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma(1+\beta2) \times \beta \times 2\left(\frac{\beta-1}{2}\right)}\right) \qquad (20)$$

Where $\Gamma(X) = (X-1)!$

3.4 Proposed RDAVA algorithm

The AVOA algorithm has the ultimate exploration and exploitation strategy; however, it does not ensure premature convergence because of an imbalance in the exploration and exploitation phases. It is considered a significant difficulty of AVOA; therefore, a novel algorithm is developed based on the combination of RDA and AVOA where the RDA strategy helps to enhance the performance of AVOA. Specifically, the RDA's strategy is to determine the best neighbor vulture and use complete historical vulture information to obtain a better optimal solution.
As discussed, the AVOA suddenly fell into local optima due to an insignificant mechanism to maintain individual historical data of vultures. Equation (7) is executed based on the current optimal solution, i.e., utilized only when the parameter $p1 \geq rand_{p1}^t$. However, the AVOA-based approaches are not as effective in exploring unknown areas. Thus, the process may lead to poor performance in the exploitation process. Therefore, an RDAVA has emerged to

maintain the record of the historical optimal solution and update the location in exploration accordingly. Thus, the exploration performance of RDAVA is much higher than the traditional AVOA. The RDAVA is executed on AVOA's exploration phase only if $rand_{p1}^t \leq p1$. In this case, the RD's strategy is added with AVOA by improving the equation (9)

$$D_i^t = \left(\sum_{j \in J}(C \times R_i^t - X_i^t)^2\right)^{1/2} \qquad (21)$$

The formulation of RDAVA is based on considering the mating strategy with the nearest hind by the stag. Using the equation below, the RDA [34] is an executed strategy for the stag to determine the nearest hind from the j-dimension.

$$d_i = \left(\sum_{j \in J}(stag_j - hind_j^i)^2\right)^{1/2} \qquad (22)$$

Where d_i to determine the distance between the current i_{th} hind and a stag. Here, based on the minimum value, the hind is selected. Similarly, based on equation (21) best optimal vulture is obtained, and its current information is stored for further processing. Therefore, an optimal solution can be obtained and balance exploration and exploitation by avoiding premature convergence. The proposed RDAVA can help select significant features and features further transferred into the two-level architecture with BiLSTM to process high-level classification. Table 2 shows the pseudocode of the RDAVA algorithm with TF-IDF.

Table 2 shows the pseudocode of the RDAVA algorithm to improve TF-IDF

INPUT: N, D, T, TF-TDF
FITNESS VALUE : $MAD_{xi} = \frac{1}{a_i}\sum_{j=1}^{t}\|x_{ij} - \hat{x}_i\|$
POPULATION INITIALIZATION: $X_i^t = x_{i1}^t, \ldots \ldots, x_{id}^t, \ldots \ldots, x_{iD}^t$
TERMINATION CRITERIA: while(stopping condition met) do
Fitness evaluation
Find the best first solution
Find the best second solution
For(each vulture (x_i))
Calculate R(i) using $R_i^t = \begin{matrix} BV_1^t, p_i^t = L_1 \\ BV_2^t, p_i^t = L_2 \end{matrix}$
Use equation $F_i^t = (2 \times rand_{i1}^t + 1) \times z^t \times \left(1 - \frac{t}{T}\right) + g^t$ to update the f
Red optimization to determine the distance
Use $D_i^t = \left(\sum_{j \in J}(C \times R_i^t - X_i^t)^2\right)^{1/2}$ *for* distance calculation
If(|f|≥1)) executes   if(p1≥randp1) executes update the solution using eq. $X_i^{t+1} = \begin{cases} R_i^t - D_i^t \times F_i^t, p_1 \geq rand_{p1}^t \\ R_i^t - F_i^t + rand_{i2}^t \times ((ub - lb) \times (rand_{i3} + lb)), p1 < rand_{p1}^t \end{cases}$ $X_i^{t+1} = R_i^t - D_i^t \times F_i^t$
Otherwise
Update the location vulture using $X_i^{t+1} = R_i^t - F_i^t + rand_{i2}^t \times ((ub - lb) \times (rand_{i3} + lb))$
$if(\|F\| < 1)$ executes $if(\|F\| \geq 0.5)$ executes $if(p_2 \geq rand_{p2})$ then   update the position using $X_i^{t+1} = R_i^t - (S_{i1}^t + S_{i2}^t)$ Otherwise   update the position using $S_{i2}^t = R_i^t \times \left(\frac{rand_6^t \times X_i^t}{2\pi}\right) \times \sin(X_i^t)$

Else if(|F| < 0.5)   If($p_3 \geq rand_{p3}$) executes   update the position using $A_{i2}^t = BV_2^t - \frac{BV_2^t \times X_i^t}{BV_2^t \times (X_i^t)^2} \times F_t^t$ Otherwise   update the position using $Levy(dim) = 0.01 \times \frac{r_1 \times \sigma}{\|r2\|^{\frac{1}{\beta}}}$
Return the final best solution
The best solution for improving the TF-IDF
Return best features

3.5 Two-level Classification with Bi-LSTM

The proposed two-level classification has incorporated RNN and RBM algorithms. The extracted features are transferred into a two-level architecture, where feature vectors are initially moved into the RNN and RBM architecture. Then the processed parts from RNN and RBM are transferred into the Optimized-Bi-LSTM for classification.

RNN-based Feature Extraction

The time series data is based on interconnected flow where individual predictions may fall into false results due to inefficient models dealing with long-term dependencies. In this scenario, the RNN provides better prediction by learning the model weights from the current cell to the last cell. It executes the reverse and forward flows, which differs the RNN from other network models, i.e., specifically, suits to Feedforward Neural Network (FNN). In RNN, the hidden states in the architecture are updated by utilizing the non-linear activation function for predicting the output. Therefore, the weight is multiplied and retransferred to the input of neurons. The preceding timestamp significantly impacts the current timestamp, which helps the process of memory allocation processes. Moreover, the memory helps to monitor the preceding computed outcome. As discussed, the RNN provides a better platform to predict future data from previous data. The last timestamp of the RNN is denoted as "t-1," which indicates the flow towards the hidden layer to the subsequent timestamp as "t." The prediction model is weights and bias. The weight and bias of the hidden layers are multiplied to predict an expected outcome by applying an appropriate activation function.

For instance, the sample time series-based input is represented as $s = s_1, s_2, s_3, \ldots.., s_T$. The S is computed by the RNN model, where the hidden layers are represented as $r_h = r_{h1}, r_{h3}, r_{h3}, \ldots. r_{hT}$. Based on computing the S and r_h, the output will be emerged as $O_p = O_{p1}, O_{p2}, O_{p3}, \ldots.., O_{pT}$. The mathematical derivation of the RNN-based computation process is shown below.

$$h_t = f\left(w_{r_{hs}}s_t + w_{r_h r_h}r_{h_{t-1}} + b_h\right) \qquad (23)$$

$$y_t = g\left(w_{O_p r_h}r_{h_t} + b_{O_p}\right) \qquad (24)$$

Where, $w_{r_{hs}}$ Denotes weight matrix of the hidden input state. $w_{r_h r_h}$ and $w_{O_p r_h}$ respectively illustrating the weight matrix of

the hidden-hidden and the hidden output states. The bias of both hidden and output are respectively denote b_{r_h} and b_{O_p}. The activation functions of the hidden layers are represented as $f(.)$, and the output layer is denoted as $g(.)$. The r_{h_t} Helps memorize the network's hidden state at the "t" time step. In addition, the r_{h_t} helps to capture the previous information of the last time steps. However, the RNN network has a limitation in long-term sequential data. The proposed model utilizes RNN-based feature extraction. The RNN selects the significant sentence from the summary.

This model has incorporated the input vectors, the layers of RNN, and summary generation. For effective sentence summarization, the given text is converted into numeric values, i.e., vector representation. The feature extraction is achieved by data received from the pre-processed data. Then the concatenated significant features using the $\oplus$ concatenated operator. For instance, the S_n is a sentence with an "n" number of sentences where all the sub-sentences are added using a concatenate operator.

Moreover, the textual feature is padded by 0. Then the RNN network processes the input feature vectors to generate the quality-enhanced summary. Furthermore, the sentence vectors are transferred into the hidden layer, where learned weights and biases are utilized to multiply the feature vectors separately. The RNN executes the back-propagation algorithm that helps to append features into the model and remove unnecessary features.

RBM-based Feature Extraction

RBM is an energy-based two-layer model with the joint probability function that has incorporated with the visible unit as $v_u = \{v_{u1}, v_{u2}, \ldots, v_{nT}\}$ and v_{nT} represents the overall visible layers and $h_u = h_{u1}, h_{u2}, \ldots, h_{uM}$ and h_{uM}, denotes the overall hidden layers. In addition, each unit in the model is connected to an opposing layer, yet there are direct linkages between variables within the same layer.

$$p(v_u, h_u) = \frac{1}{P_f} e^{-E(v_u, h_u)}, \qquad (25)$$

$$P_f = \sum_{v_u} \sum_{h_u} e^{-E(v_u, h_u)} \qquad (26)$$

Where P_f As the partition function, i.e., utilized as the constant for normalization. The following shows the mathematical derivation of the energy function

$$E(v_u, h_u) = -b_v^T v_u - c_h^T h_u - h_u^T w_v \qquad (27)$$

Where, b_v and c_h are respectively denote the bias of visible and hidden layers. W as a weight matrix that incorporated the elements as w_{ij}. As discussed earlier, the visible and hidden units are independently executed based on specific conditions. The conditional probability is calculated based on energy function and joint distribution. The following equation shows the probabilistic function based on the activation function.

$$p\left(h_{uj} = 1|v_u\right) = \sigma\left(b_{v_j} + \sum_i v_{ui} W_{ij}\right)$$

$$p\left(v_{uj} = 1|h_u\right) = \sigma\left(a_{v_j} + \sum_j W_{ij} h_{uj}\right) \qquad (28)$$

Where $\sigma(x) = \frac{1}{1+e^x}$ It is the sigmoid activation function. Similar to RNN, the features obtained by RBM are fed into the Optimized BiLSTM.

3.6 Bi-LSTM based classification

The LSTM model is highly effective for processing long-term dependencies that RNNs can achieve. The behavior of LSTM is similar to a conveyor belt because its process flow on the cell state is straight down. The LSTM will not execute any add or remove operation within a cell. In the architecture of the LSTM, the network mode has several gates that are incorporated with pointwise multiplication operators and sigmoid functions. The first step is to throw the actual information using LSTM. The decision-making process is done by executing the sigmoid layer. The information passed via the forget gate is represented as ht−1 and xt. Therefore, two possibilities of results can be observed, such as 0 or 1. The overall operation related to the flow of information is based on several sigmoid values, and the mathematical formulation is shown below.

$$\begin{aligned}
ft &= \sigma(wf.[h_{t-1}, xt] + bf) \\
i_t &= \sigma(w_i.[ht-1, xt] + b_i) \\
C_t' &= tanh(w_c.[ht-1, xt] + b_i) \\
C_t &= f_t * C_{t-1} + i_t * C_t' \\
o_t &= \sigma(W_o.[ht-1, xt] + b_o)
\end{aligned} \qquad (29)$$

where, ft = Forget state, it = Input state, C′t = Intermediate state, Ct = Cell state and ot = Output state. The complete output of the LSTM operation is based on the cell state with filtered output. The BiLSTM receives the feature vectors from the RNN and the RBM for classifying fake news. The received features are represented as $x_i = x1, x2, \cdots\cdots x_n$, in which i denotes the current word in the sentence. In the BiLSTM architecture, the forward propagation illustrates data flow in forward order that can be represented as x1 → xn. Similarly, the data flow in the backward order is described as xn → x1. The forward and backward propagation are significant factors of the BiLSTM as the traditional LSTM can only execute the data transmission from front to backflow, the less possibility of maintaining the context information in all the cells. Moreover, the BiLSTM can facilitate the network model to receive the current work with the semantic information by ensuring the context information in all the cells.

The feature vectors obtained from RNN and RBM are fused to achieve high-level classification performance. The obtained features are combined by applying the concatenation operator $\oplus$. For instance, let us consider the RNN-based feature vectors as $R_v = R_{v1}, R_{v2}, \ldots\ldots, R_{vn}$ and RBM feature as $R_b = R_{b1}, R_{b2}, \ldots\ldots, R_m$. The fused vectors $F_{vec} = R_v \oplus R_b$ Are transferred into the fully connected layer with the Softmax activation function. The probability distribution-based mathematical calculation is executed as follows.

$$p(y = j|F_{vec}) = \frac{e^{F_{vec}}}{\sum_{k=1}^{k} F_{vec_l} w_l} \quad (30)$$

Where, F_{vec} Represents the sample vector, k represents the number of classes, the weight vector is w, and the label is j. Moreover, the RDAVA algorithm applies in the LSTM to achieve the optimized outcome. Therefore, the parameters such as minibatch, epoch, momentum, etc., are considered and executed in the hyperparameters tuning operation. The cross entropy-based loss function is deemed to improve model performance. The mathematical form of the loss function is shown below.

$$Loss = -\sum_{n=1}^{B} y_i \log(\widehat{y_i}) \quad (31)$$

Where B represents the class label, y_i denotes the class at the i_{th} iteration, and the predicted outcome as $\widehat{y_i}$. Moreover, the Dropout operation can be performed to prevent over-fitting issues. Eventually, the classification results are obtained in terms of fake or real news.

Table 3: Pseudocode of RDAVA in BiLSTM

INITIALIZATION N, D, T, epoch, mini-batch, momentum, etc. of RDAVA and BiLSTM
FITNESS EVALUATION $Cross\ Entropy\ based\ Loss = -\sum_{n=1}^{B} y_i \log(\widehat{y_i})$ $if\left((\|F\| \geq 1)\,\delta\left(p_1 \geq rand_{p1}\right)\right)$ EQ.(6) $X_i^{t+1} = R_i^t - D_i^t \times F_i^t$ to update the position of vultures

Otherwise
 Execute the Second condition in Eq.(6) to Update the location vulture
  $X_i^{t+1} = R_i^t - F_i^t + rand_{i2}^t \times ((ub - lb) \times (rand_{i3}^t + lb))$
$if\left((\|F\| < 1)\,\delta\,(\|F\| \geq 1)\delta\left(p_2 \geq rand_{p2}\right)\right)$
 Execute Eq.(8) for updating the position
Otherwise
 Eq.(10) to update
Else if$(\|F\| < 0.5)$
  if$(p_3 \geq rand_{p3})$
Executes the updation using Eq. (15)
Otherwise
 Equ. (18) to position update
Return Best solutions to BiLSTM

Classification accuracy will be improved

When RDAVA is included, Bi-LSTM models perform noticeably better on a variety of tasks, including time series prediction, sentiment analysis, and sequence labelling. RDAVA enhances feature selection through the efficient identification of pertinent features, reduction of dimensionality, and improvement of the model's capacity to capture critical patterns. Furthermore, it facilitates the fine-tuning of hyperparameters, thereby addressing concerns such as overfitting and underfitting, and optimizes the

training procedure by providing direction in seeking optimal weights and biases. The combined utilization of RDAVA and Bi-LSTM models results in accelerated convergence, enhanced solution quality, and improved solution quality.

4. Result and Discussion

This section describes the proposed model's performance based on performance metrics, including accuracy, specificity, sensitivity, F-measure, and precision. In addition, the error metrics are comprised of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These metrics evaluate the proposed model using ISOT, BuzzFeed, and FakeNewsNet Datasets. Each dataset is used individually with the proposed classifier to assess the performance of Feature Selection (FS) with RDAVA and 2-level classification with RDAVA-optimized Bi-LSTM.

In addition, Table 3 contains the comparison tables for feature selection with and without the RDAVA algorithm on BuzzFeed, FakeNewsNet, and ISOT DSs. This classification aids in determining the effect of improved FS on classification performance. In this analysis, the RDAVA algorithm achieves a higher ACC on BuzzFeed and ISOT DS, obtaining 98% and 98%, respectively. In contrast, the FakeNewsNet DS utilizing the RDAVA algorithm attained 97% accuracy, which is lower than the other datasets.

The performance of 2-level classification with RDAVA-based Optimized Bi-LSTM (Proposed algorithm) is compared to other techniques, including Bi-LSTM, RNN+RBM, RNN, RBM, and ANN. Similar to feature selection-based evaluation, the classification performance is analyzed here. Consequently, we have considered error-measure-based parameters, including TP, TN, FP, and FN. It is intended that datasets including BuzzFeed, FakeNewsNet, and ISOT were employed on the proposed algorithm for evaluating classification performance. Based on three distinct datasets, the performance of the proposed algorithm is assessed, and the results are tabulated separately to demonstrate that the proposed model is vastly superior to other comparable models. The comparative results of the proposed algorithm were applied on ISOT, Buzzfeed and FakeNewsNet and results are analyzed in table 4, 5 and 6 respectively.

Table 4: Comparative result of ISOT-based evaluation

Techniques	Accuracy	Precision	F-Measure
Proposed algorithm	0.9863	0.9898	0.9862
Bi-LSTM	0.9729	0.9839	0.9736
RNN + RBM	0.9347	0.9187	0.9386
RNN	0.8990	0.9121	0.9019
RBM	0.8846	0.9253	0.8841
ANN	0.8613	0.8837	0.8636

Table 3: Comparative result based for feature selection (with RDAVA and without RDAVA)

BuzzFeed					
Techniques	Sensitivity	Specificity	Accuracy	Precision	F-Measure
With RDAVA	0.9762	1	0.9881	1	0.9880
Without RDAVA	0.7381	0.6905	0.7143	0.7045	0.7209
FakeNewsNet DS					
With RDAVA	0.9690	0.9747	0.9721	0.9708	0.9699
Without RDAVA	0.8553	0.8718	0.8642	0.8530	0.8542
ISOT DS					
With RDAVA	0.9838	0.9890	0.9863	0.9898	0.9868
Without RDAVA	0.7969	0.8076	0.8020	0.8178	0.8072

Table 5: Comparative result of BuzzFeed-based evaluation

Techniques	Accuracy	Precision	F-Measure
Proposed algorithm	0.9881	1	0.9880
Bi-LSTM	0.9643	0.9756	0.9639
RNN + RBM	0.9048	0.8864	0.9070
RNN	0.8929	0.8837	0.8941
RBM	0.7857	0.7608	0.7955
ANN	0.7619	0.7391	0.7727

Table 6: Comparative result of FakeNewsNet-based evaluation

Techniques	Accuracy	Precision	F-Measure
Proposed algorithm	0.9721	0.9708	0.9699
Bi-LSTM	0.9537	0.9554	0.9499
RNN + RBM	0.9485	0.9405	0.9449
RNN	0.9236	0.9182	0.9178
RBM	0.9100	0.9126	0.9022
ANN	0.8960	0.8854	0.8887

The proposed method works very well for the three datasets selected. Figure 2 shows the detailed accuracy, precision and F-measure comparison of proposed methods on FakeNewsNet, Buzzfeed and ISOT datasets. This comparative analysis proves the proposed model performance as a highly efficient model that is better than all the other existing approaches.
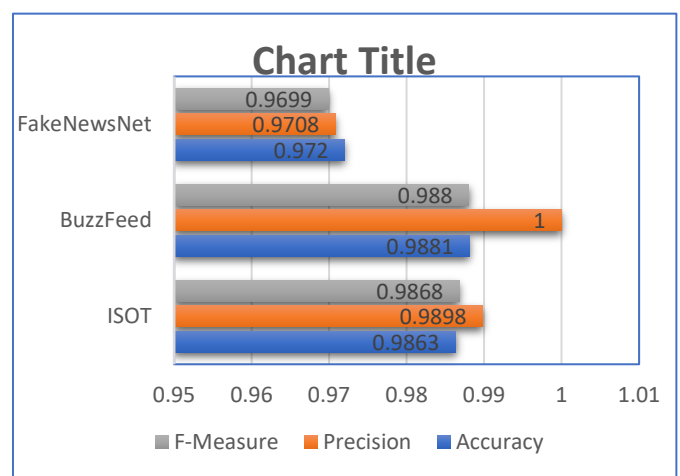

Fig. 2 Overall performance metric measures of the proposed model

Conclusion and Future Scope

The proposed algorithm for detecting fake news on online social media sites has been effectively implemented. The successful model can be determined by analyzing multiple datasets; thus, the proposed work was conducted on BuzzFeed, ISOT, and FakeNewsNet. The proposed RDAVA algorithm can provide high-level feature selection by acquiring significant features. The comparison is therefore performed with and without RDAVA-based feature selection. The feature selection from the BuzzFeed dataset revealed that the accuracy is 0.9880 with RDAVA and 0.7143 without it. Then, the FakeNewsNet dataset-based evaluation demonstrated that RDAVA achieved 0.9721 accuracies, whereas it achieved 0.8642 without RDAVA.

Similarly, the ISOT dataset obtained 0.9863 accuracies with RDAVA but only 0.802017 without RDAVA. Second, a 2-level classification is implemented in which RNN and RBM are separately processed for feature extraction and then transmitted to the Bi-LSTM model. In addition, the proposed model employs a highly optimized

Bi-LSTM to achieve superior classification performance. To enhance the Bi-LSTM-based classification, the RDAVA is also used for hyper-parameter tuning. Fortunately, the proposed algorithm performs well in fake news prediction, achieving 98% accuracy on the BuzzFeed and ISOT datasets but only 97% accuracy on the FakeNewsNet dataset. Two-level analyses, such as feature selection and classification, help demonstrate this point. In the BuzzFeed and ISOT datasets, the RDAVA algorithm performs superiorly for feature selection. Similarly, classification results incorporate RDAVA with two-level classification and Optimized Bi-LSTM more effectively.

The generalizability of the suggested system may be restricted due to its dependence on datasets for both training and assessment. Although it achieves good accuracy on selected datasets (BuzzFeed, ISOT, FakeNewsNet), its performance may differ when applied to other datasets from various sources or datasets with different characteristics.

The proposed model will be enhanced by considering the multimodal framework in future work. Moreover, the feature selection model's efficiency will be further enhanced by incorporating a high-level feature selection approach. Future work will be improved by executing a high-level comprehensive performance evaluation and eventually considering the best result for detecting fake news more effectively.

References

[1] H. Saleh, A. Alharbi and S. H. Alsamhi, "OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection," in *IEEE Access*, vol. 9, pp. 129471-129489, 2021. doi: 10.1109/ACCESS.2021.3112806

[2] M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar and M. S. Rahman, "A Comprehensive Review on Fake News Detection With Deep Learning," in *IEEE Access*, vol. 9, pp. 156151-156170, 2021. doi: 10.1109/ACCESS.2021.3129329

[3] T. Jiang, J. P. Li, A. U. Haq, A. Saboor and A. Ali, "A Novel Stacking Approach for Accurate Detection of Fake News," in *IEEE Access*, vol. 9, pp. 22626-22639, 2021. doi: 10.1109/ACCESS.2021.3056079

[4] D. Rohera *et al.*, "A Taxonomy of Fake News Classification Techniques: Survey and Implementation Aspects," in *IEEE Access*, vol. 10, pp. 30367-30394, 2022. doi: 10.1109/ACCESS.2022.3159651

[5] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi and B. -W. On, "Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM)," in *IEEE Access*, vol. 8, pp. 156695-156706, 2020. doi: 10.1109/ACCESS.2020.3019735

[6] H. Choi and Y. Ko, "Using Adversarial Learning and Biterm Topic Model for an Effective Fake News Video Detection System on Heterogeneous Topics and Short Texts," in *IEEE Access*, vol. 9, pp. 164846-164853, 2021. doi: 10.1109/ACCESS.2021.3122978

[7] W. Shishah, "JointBert for Detecting Arabic Fake News," in *IEEE Access*, vol. 10, pp. 71951-71960, 2022. doi: 10.1109/ACCESS.2022.3185083

[8] G. Shan, B. Zhao, J. R. Clavin, H. Zhang and S. Duan, "Poligraph: Intrusion-Tolerant and Distributed Fake News Detection System," in *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 28-41, 2022. doi: 10.1109/TIFS.2021.3131026

[9] P. Wei, F. Wu, Y. Sun, H. Zhou and X. -Y. Jing, "Modality and Event Adversarial Networks for Multi-Modal Fake News Detection," in *IEEE Signal Processing Letters*, vol. 29, pp. 1382-1386, 2022. doi: 10.1109/LSP.2022.3181893

[10] P. K. Verma, P. Agrawal, I. Amorim and R. Prodan, "WELFake: Word Embedding Over Linguistic Features for Fake News Detection," in *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 881-893, Aug. 2021. doi: 10.1109/TCSS.2021.3068519

[11] K. A. Qureshi, R. A. S. Malick, M. Sabih and H. Cherifi, "Complex Network and Source Inspired COVID-19 Fake News Classification on Twitter," in *IEEE Access*, vol. 9, pp. 139636-139656, 2021. doi: 10.1109/ACCESS.2021.3119404

[12] Choudhury, D. and Acharjee, T., 2022. A novel approach to fake news detection in social networks using genetic algorithm applying machine-learning classifiers. Multimedia Tools and Applications, pp.1-17.

[13] Meesad, P., 2021. Thai fake news detection based on information retrieval, natural language processing and machine learning. SN Computer Science, 2(6), pp.1-17.

[14] Taskin, S.G., Kucuksille, E.U. and Topal, K., 2022. Detection of Turkish Fake News in Twitter with Machine Learning Algorithms. Arabian Journal for Science and Engineering, 47(2), pp.2359-2379.

[15] Braşoveanu, A.M. and Andonie, R., 2021. Integrating machine-learning techniques in semantic fake news detection. Neural Processing Letters, 53(5), pp.3055-3072.

[16] Raza, S. and Ding, C., 2022. Fake news detection based on news content and social contexts: a transformer-based approach. International Journal of Data Science and Analytics, 13(4), pp.335-362.

[17] Uppada, S.K., Manasa, K., Vidhathri, B., Harini, R. and Sivaselvan, B., 2022. Novel approaches to fake news and fake account detection in OSNs: user social engagement and visual content centric model. Social Network Analysis and Mining, 12(1), pp.1-19.

[18] Palani, B., Elango, S. and Viswanathan K, V., 2022. CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT. Multimedia Tools and Applications, 81(4), pp.5587-5620.

[19] de Souza, M.C., Nogueira, B.M., Rossi, R.G., Marcacini, R.M., Dos Santos, B.N. and Rezende, S.O., 2021. A network-based positive and unlabeled learning approach for fake news detection. Machine Learning, pp.1-44.

[20] Sadeghi, F., Bidgoly, A.J. and Amirkhani, H., 2022. Fake news detection on social media using a natural language inference approach. Multimedia Tools and Applications, pp.1-21.

[21] Kaliyar, R.K., Goswami, A. and Narang, P., 2021. DeepFakE: improving fake news detection using tensor decomposition-based deep neural network. The Journal of Supercomputing, 77(2), pp.1015-1037.

[22] Choudhary, A. and Arora, A., 2021. Linguistic feature based learning model for fake news detection and classification. Expert Systems with Applications, 169, p.114171.

[23] Nasir, J.A., Khan, O.S. and Varlamis, I., 2021. Fake news detection: A hybrid CNN-RNN based deep learning approach. International Journal of Information Management Data Insights, 1(1), p.100007.

[24] Kumar, S., Asthana, R., Upadhyay, S., Upreti, N. and Akbar, M., 2020. Fake news detection using deep learning models: A novel approach. Transactions on Emerging Telecommunications Technologies, 31(2), p.e3767.

[25] Ni, S., Li, J. and Kao, H.Y., 2021. MVAN: Multi-View Attention Networks for Fake News Detection on Social Media. IEEE Access, 9, pp.106907-106917.

[26] Kaliyar, R.K., Goswami, A. and Narang, P., 2021. EchoFakeD: improving fake news detection in social media with an efficient deep neural network. Neural computing and applications, 33(14), pp.8597-8613.

[27] Kumari, R. and Ekbal, A., 2021. Amfb: attention based multimodal factorized bilinear pooling for multimodal fake news detection. Expert Systems with Applications, 184, p.115412.

[28] Sharma, D.K. and Garg, S., 2021. IFND: a benchmark dataset for fake news detection. Complex & Intelligent Systems, pp.1-21.

[29] Yuan, H., Zheng, J., Ye, Q., Qian, Y. and Zhang, Y., 2021. Improving fake news detection with domain-adversarial and graph-attention neural network. Decision Support Systems, 151, p.113633.

[30] Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G.S. and On, B.W., 2020. Fake news stance detection using deep learning architecture (CNN-LSTM). IEEE Access, 8, pp.156695-156706.

[31] Sansonetti, G., Gasparetti, F., D'aniello, G. and Micarelli, A., 2020. Unreliable users detection in social media: Deep learning techniques for automatic detection. IEEE Access, 8, pp.213154-213167.

[32] Han, B., Han, X., Zhang, H., Li, J. and Cao, X., 2021. Fighting fake news: two stream network for deepfake detection via learnable SRM. IEEE Transactions on Biometrics, Behavior, and Identity Science, 3(3), pp.320-331.

[33] Amir Mohammad Fathollahi-Fard, Mostafa Hajiaghaei-Keshteli & Reza Tavakkoli-Moghaddam , "Red deer algorithm (RDA): a new nature-inspired meta-heuristic", Soft Computing, Vol.24, 2020

[34] Benyamin Abdollahzadeh,Farhad Soleimanian Gharehchopogh,Seyedali Mirjalili, "African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems", Computers & Industrial Engineering, 2021.

[35] https://www.kaggle.com/code/sohamohajeri/buzzfeed-news-analysis-and-classification/data

[36] https://www.uvic.ca/ecs/ece/isot/datasets/fake-news/index.php

[37] https://github.com/KaiDMML/FakeNewsNet

[38] Chandraprabha M, Dhanraj RK. Ensemble Deep Learning Algorithm for Forecasting of Rice Crop Yield based on Soil Nutrition Levels. EAI Endorsed Transactions on Scalable Information Systems. 2023 May 4;10(4).

[39] Singh R, Subramani S, Du J, Zhang Y, Wang H, Miao Y, Ahmed K. Antisocial Behavior Identification from Twitter Feeds Using Traditional Machine Learning Algorithms and Deep Learning. EAI Endorsed Transactions on Scalable Information Systems. 2023 May 12;10(4).

[40] Ge YF, Wang H, Bertino E, Zhan ZH, Cao J, Zhang Y, Zhang J. Evolutionary dynamic database partitioning optimization for privacy and utility. IEEE Transactions on Dependable and Secure Computing. 2023 Aug 7.

[41] Yang JQ, Yang QT, Du KJ, Chen CH, Wang H, Jeon SW, Zhang J, Zhan ZH. Bi-directional feature fixation-based particle swarm optimization for large-scale feature selection. IEEE Transactions on Big Data. 2022 Dec 29.

[42] Koroteev MV. BERT: a review of applications in natural language processing and understanding. arXiv preprint arXiv:2103.11943. 2021 Mar 22.

[43] González-Carvajal S, Garrido-Merchán EC. Comparing BERT against traditional machine learning text classification. arXiv preprint arXiv:2005.13012. 2020 May 26.

[44] Stewart J, Lyubashenko N, Stefanek G. The efficacy of detecting AI-generated fake news using transfer learning. Issues in Information Systems. 2023 Apr 1;24(2).

[45] Oshikawa R, Qian J, Wang WY. A survey on natural language processing for fake news detection. arXiv preprint arXiv:1811.00770. 2018 Nov 2.

[46] Kapusta J, Hájek P, Munk M, Benko Ľ. Comparison of fake and real news based on morphological analysis. Procedia Computer Science. 2020 Jan 1;171:2285-93.

[47] Fard AF, Hajiaghaei-Keshteli M. Red Deer Algorithm (RDA); a new optimization algorithm inspired by Red Deers' mating. InInternational Conference on Industrial Engineering, IEEE 2016 Dec (Vol. 12, pp. 331-342).

[48] Bovet A, Makse HA. Influence of fake news in Twitter during the 2016 US presidential election. Nature communications. 2019 Jan 2;10(1):7.

[49] Rocha YM, de Moura GA, Desidério GA, de Oliveira CH, Lourenço FD, de Figueiredo Nicolete LD. The impact of fake news on social media and its influence on health during the COVID-19 pandemic: A systematic review. Journal of Public Health. 2021 Oct 9:1-0.