MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2138-2152

LATIN HYPERCUBE SAMPLING HIPPOPOTAMUS OPTIMIZATION (LHSHO) BASED FEATURE SELECTION AND STACKING ENSEMBLE DEEP LEARNING (SEDL) MODEL FOR FAKE NEWS DETECTION

Mrs. L.Padmavathy
Research Scholar, Department of Computer Science,
PSGR Krishnammal college for women,
padmavathy2988@gmail.com
Dr. S.Nithya
Assistant Professor, Department of Computer Science,
PSGR Krishnammal college for women,
nethanithashini@gmail.com

ABSTRACT: Fake news's online development through platforms has become an important concern, affecting political stability, public opinion, and the spread of trustworthy information. Machine learning (ML) and deep learning (DL) techniques are introduced to identify and classify fake news. Several features are presented in the detection model, features are manually extracted from news, and the input text includes lengthy paragraphs. It becomes reduces the detection accuracy of the model. To solve this issue, in this paper novel summarization, and fake news detection method is introduced for social networks. Firstly, a fake news dataset is gathered from BuzzFeed and PolitiFact. Secondly, Hybrid Bidirectional Encoder Representation from Transformers (HBERT) is developed for text summarization to consider both fine-grained text information and label semantics at an identical period. Thirdly, text preprocessing and feature extraction are performed for summarized text. Fourthly, Latin Hypercube Sampling Hippopotamus Optimization (LHSHO) is introduced for feature selection which includes the hippopotamus natural reaction of running away from predators and dynamically attempting to distance itself from reduced error. LHS will help to prevent the initial population generation, and solves local optimal problem during selection process. Lastly, the Stacking Ensemble Deep Learning (SEDL) model is introduced, which combines various models such as Conditional Generative Adversarial Network (CGAN), Self Adaptive Attention-based bidirectional Long Short-Term Memory (SAA-BiLSTM), and Stacked Sparse Auto Encoder (SSAE). Projection-based extension of GAN is called CGAN. SSAE, a sparse penalty term is employed to achieve more accurate and effective detection. Stacking is used to combine the results of these classifiers by training a meta-model (Logistic Regression (LR)). Performance is assessed using metrics including precision, recall, F-measure, and accuracy.

KEYWORDS: Social media, fake news, feature selection, Text summarization, Hybrid Bidirectional Encoder Representation from Transformers (HBERT), Latin Hypercube Sampling Hippopotamus Optimization (LHSHO), Stacking Ensemble Deep Learning (SEDL), Conditional Generative Adversarial Network (CGAN), and Stacked Sparse Auto Encoder (SSAE).

1. INTRODUCTION

Facebook, Instagram, WhatsApp, and Twitter have expanded at an incredible pace over the past ten years, exceeding traditional print media and becoming the main avenues for spreading and sharing of news [1]. People are increasingly using online social networks (OSNs) as their primary source of daily news and information as a result of developing social interaction patterns. At any time, users can use the internet to communicate with other network members [2]. These OSNs have also worked as platforms for people to express their thoughts, feelings, and concerns while obtaining quick responses and comments on local or global issues. Since social networks may distribute news far more quickly than traditional media, they are shown to be more beneficial [3].

Since both fake and real information spread quickly, OSN's replacement of traditional media has an even more rapid effect [4]. The spread of fake news has a negative effect on its intended target population as well as society at large. Therefore, it affects public opinion and reactions to real news, disturbing the balance of the news system [4-5]. By affecting people's thoughts and emotions [6], changing their views, and possibly leading to negative actions [7], fake news has an effect on people's daily lives. Social media's spread of false information has a negative impact on society in a variety of domains, including politics, the economy, society, health, technology, and sports [8].

OSN platforms offer a virtual space where users can post, converse, share opinions, and engage globally without being constrained by time, place, or volume of material. According to recent studies, the proliferation of fake news on OSNs has escalated into a pressing worldwide issue that has to be promptly addressed and curbed since it can cause social unrest and even economic instability [9]. The rapid growth of OSNs and people's thirst for real-time news make it crucial to identify the news's veracity as soon as possible. Given these conditions, identifying fake news quickly and accurately has become a crucial problem that OSNs must address immediately. Numerous researches on fake news detection rely on one or more characteristics, like user behavior, network propagation, or content [10–11]. Current learning methods for detecting fake news can be broadly divided into two categories: (1) news content-based learning, which focuses primarily on news articles, the manner in which malicious news is written, linguistic styles, etc. [12]; and (2) social context-based learning, which focuses primarily on user-related behaviors (such as user type, comments, likes, shares, etc.). [13–14]. Although the research community has devoted considerable attention to fake news detection, there is still a need for an efficient model that can simultaneously handle context, community-level information, and content through a tensor-factorization approach.Identifying fake news on social media presents a number of difficulties. First of all, it is exceedingly challenging to identify fake news in input texts since they contain lengthy sequences. Second, an excessive number of features affect the performance and efficiency of model. Currently, a lot of fake-news recognition systems use attributes that are manually taken from news and then used to categorize them. Text summarization reduces lengthy sequences, Swarm Intelligence (SI) techniques remove too many features, and artificial intelligence (AI) is developed to automatically identify false information.Extractive and abstractive methods are the two categories into which text summarization falls [15–16]. Abstractive summarization generates summaries by interpreting the semantic meaning of the input text, while extractive summarization constructs a summary by selecting and combining the most important pieces of information directly from the input [17]. Coherence, ambiguity, and redundancy are issues with extraction-based summaries, even though they are syntactically correct [18–19]. To overcome the limitations of the extractive method, the abstractive approach emphasizes understanding and representing the text's underlying semantics. Word representations based on the Bidirectional Encoder Representations from Transformers (BERT) have demonstrated some extremely promising results in the field of summarization performance, and because of parallelization, it trains quickly.

Non-deterministic polynomial time (NP-hard) computational problems have been demonstrated to be solved by SI. It has been effectively applied to FS in certain applications and is becoming more and more well-liked for resolving various optimization issues. A population of simple agents interacting with their surroundings and one another in close proximity makes up a SI system. Natural systems, particularly biological ones, are a common source of inspiration. The actions of hippopotamuses served as the inspiration for the nature-inspired optimization method known as Hippopotamus Optimization (HO) [20–21].As artificial intelligence (AI) has advanced, numerous researchers have tried to use AI to automatically identify false news. Recently, Deep Learning (DL) has emerged as a promising tool in the research community and has shown greater efficiency in detecting fake news. A) automated feature extraction; b) little reliance on data pre-processing; c) the capacity to extract high-dimensional features; and d) improved accuracy are some of the specific benefits that DL offers over ML. Convolutional Neural Network (CNN), Deep Neural Network (DNN), and Long Short-Term Memory (LSTM) are most famous methods in DL. When working with complicated datasets or when individual models are prone to overfitting, ensemble approaches can frequently achieve higher accuracy than individual models [22–23]. In this paper, new feature selection and deep ensemble classification model for detecting fake news. Latin Hypercube Sampling Hippopotamus Optimization (LHSHO) was developed to eliminate irrelevant words from the text. The LHSHO algorithm selects the best features from the dataset by having hippopotamuses rotate toward the predator and their inadequate mouth. For HO, Latin hypercube sampling (LHS) is introduced for the initialization population generation. Stacking is an ensemble strategy that combines different methods, like Conditional Generative Adversarial Network (CGAN), SAA-BiLSTM, and Stacked Sparse Auto Encoder (SSAE).To improve detection accuracy, the Adam optimizer was added, and a binary loss function was added for detection. It has tested detection techniques against performance evaluation measures using real-world datasets.

2. LITERATURE REVIEW

Elsaeed et al. [25] proposed an innovative method for fake news detection that integrates voting classifiers with feature extraction and selection techniques. The process begins with dataset preprocessing, where lemmatization is applied to minimize dictionary size and remove unnecessary characters or digits. Next, various feature extraction approaches are employed, including document-to-vector representation, term frequency–inverse document frequency (TF-IDF), and word embeddings. For feature selection, the authors used Analysis of Variance (ANOVA) and Chi-Square (CS) tests to identify the most relevant features. Finally, multiple voting classifiers were implemented for classification, combining models such as Naïve Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Stochastic Gradient Descent (SGD), and Passive-Aggressive Classifier (PAC).

Jain et al. [26] presented an approach comprising feature extraction, feature selection, classification, and dataset. Linguistic Features (LFs) are extracted during the extraction process, and ambiguities from the data preparation are removed from the texts. For classification, features are chosen using the Harris Hawks Optimization (HHO) algorithm, and a hybrid model of Convolutional Neural Network-Bidirectional Long Short-Term Memory Network (CNN-BiLSTM) is used. The CNN model extracts spatial features, and BiLSTM with sequential learning capabilities improves accuracy. Interestingly, the hybrid model and HHO in the suggested approach improve the accuracy of fake news identification.

Verma et al. [27] proposed a ScrutNet, a deep learning ensemble network and distinguish between real and fake news articles effectively even with limited data. ScrutNet is launched, utilizing the cooperative strengths of a BiLSTM and a CNN. The strength of ScrutNet is its capacity for holistic understanding of text, which facilitates informed choices. The proposed ensemble classifier utilizes transformer-based models to effectively learn both sequential dependencies and local contextual patterns within textual data. To enhance performance, the parameters of individual models as well as the ensemble strategy were fine-tuned through extensive experimentation. Specifically, a grid search approach was applied to optimize critical hyperparameters, including the number of Bi-LSTM units, CNN filter and kernel sizes, and dropout rates. On the partition test of the Fake News dataset, ScrutNet offers the best detection results in terms of precision, recall, specificity, F1-score, and accuracy.

Aslam et al. [28] suggested an ensemble-based deep learning model for fake news detection. Various preprocessing techniques, such as handling missing values and normalization, were applied to the dataset. Various NLP methods, including tokenization, lemmatization, and stop word removal, were employed for preprocessing the statement attribute. The Dense Neural Network (DNN) has been proposed for the other attributes, while the Bi-LSTM and bidirectional Gated Recurrent Unit (Bi-GRU) models have been used for the statement attribute. The results of the two models (Bi-LSTM and Bi-GRU) are concatenated after applying global max and global average pooling to produce a single value. LIAR dataset, precision, recall, F-score, and accuracy of the suggested model were evaluated with existing methods.

Venkatachalam et al., [29] introduced a deep learning-based Fake News Detection (DeepFND). This method uses ensemble models from Visual Geometry Group 19 (VGG-19) and Bi-LSTM to detect false information propagated over social media. Tokenization, stemming, stop word removal, punctuation, Hypertext Markup Language (HTML) tags, Uniform Resource Locator (URL), and emotions were among the pre-processed elements used to eliminate noise from the datasets. Using a shared deep-feature extractor, the ensemble technique builds learners and uses attention modules, such as pre-training and fine-tuning, to differentiate them. This technique successfully addresses overfitting while lowering training time, memory needs, and model complexity. A unique loss function that compels the students to concentrate on particular elements of the incoming news is created by using an attention mechanism. On five datasets, many techniques for detecting fake news are shown.

Qu et al. [30] introduced a quantum multimodal fusion-based model for fake news detection (QMFND), which leverages quantum computing principles to improve efficiency and accuracy. In this approach, multimodal features are encoded into a Variational Quantum Circuit (VQC) through amplitude encoding, enabling lower computational complexity. To process high-dimensional data, the model incorporates quantum encoding along with a Quantum Convolutional Neural Network (QCNN). Experimental evaluation on the Gossip and Politifact social media datasets shows that QMFND achieves performance comparable to, or better than, conventional detection models. Furthermore, parameter optimization is carried out using a search algorithm. Notably, the QCNN circuit not only offers strong expressive and entangling capabilities but also exhibits robustness against quantum noise.

Kaliyar et al. [31] proposed a tree-based ensemble machine learning framework, specifically Gradient Boosting, for fake news detection by integrating both content-level and context-level information. The model employs gradient descent algorithms to develop adaptive boosting strategies, where key components and parameters are optimized to maximize a unified objective function. Experiments were conducted on the multi-class Fake News Challenge (FNC) dataset, applying various models for classification. The results demonstrate that the proposed ensemble framework achieves superior performance compared to existing benchmark methods.

Ali et al. [32] suggested fake news, deep ensemble model that makes use of sequential deep learning. News articles were preprocessed using NLP techniques, and their features were represented through the TF-IDF method, further enriched with n-grams. Next, an ensemble of binary classifiers, built on sequential deep learning networks, was trained to capture representative hidden features capable of distinguishing different categories of news. The outputs from these binary classifiers were then combined and used as input for a Multilayer Perceptron (MLP)-based multi-class classifier to perform the final classification. The proposed model was evaluated on two widely used benchmark datasets, LIAR and ISOT, against multiple baseline classifiers. Experimental results showed that the model achieved significant improvements in precision, recall, F1-score, and accuracy compared to state-of-the-art methods that relied on deep contextualized representations with CNNs.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2138-2152

Padalko et al. [33] proposed an ensemble machine learning model for fake news detection that combined Light Gradient Boosting Machine (LightGBM), eXtreme Gradient Boosting (XGBoost), and Balanced Random Forest (BRF). To prepare the dataset, they removed noise such as HTML tags and URLs while keeping only the meaningful text. The models were then trained on a large and diverse collection of news articles to capture different writing styles and linguistic patterns. Features were extracted using TF-IDF, which identifies the importance of words based on their frequency within and across documents. Incorporating influential tokens further improved the accuracy and reliability of the classifiers. In this framework, LightGBM built trees sequentially, where each tree corrected errors from the previous one, while BRF addressed class imbalance by adjusting distributions before training. XGBoost enhanced performance by training new models to predict the residual errors of earlier models. Overall, the ensemble achieved strong and scalable performance, showing its potential as an effective solution for fake news detection. Beyond accuracy, the study emphasized its role in strengthening information integrity, safeguarding public trust, and supporting informed discussions.

Jiang et al. [34] assessed the performance of three deep learning models (CNN, LSTM, and GRU) and five machine learning models (Decision Tree (DT), RF, Logistic Regression (LR), Support Vector Machine (SVM), and K Nearest Neighbour (KNN)) on two fake and real news datasets of varying sizes with holdout cross validation. ext representations were generated using TF and TF-IDF for machine learning models, while embedding-based methods such as CNN, LSTM, and GRU were employed for deep learning model. Precision, recall, F1-score, and accuracy are used as evaluation metrics, and a modified McNemar's test is used to assess model performance. The novel stacking model, which achieved the highest accuracy on the ISOT and KDnugget dataset was then evaluated.

Kaliyar et al. [35] introduced a deep convolutional neural network model with fake news detection (FNDNet). The network multiple hidden layers are employed to automatically learn and extract discriminative features for classification, with CNN layers capturing various feature representations at each stage. The performance of FNDNet was evaluated against several baseline models using metrics such as the Wilcoxon test, false positive, true negative, precision, recall, F1-score, and accuracy. Experimental results demonstrated significant improvements over existing state-of-the-art methods, highlighting the potential of FNDNet for effective fake news detection on social media platforms.

3. PROPOSED METHODOLOGY

Initially, dataset of fake news is gathered from a benchmark repository. Secondly, the text is summarized with HBERT, which takes into account both label semantics and fine-grained text details. Thirdly, text preprocessing and feature extraction methods are presented for the summarized text. Fourthly, irrelevant words in the text are eliminated through Latin Hypercube Sampling Hippopotamus Optimization (LHSHO). Finally, the Stacking Ensemble Deep Learning (SEDL) model is presented by combining the results of individual classifiers (CGAN, SAA-BiLSTM, and SSAE) by training a meta-model (Logistic Regression (LR)). It learns to make detections based on the outputs of the individual methods. Figure 1 shows the overall block diagram of proposed model.

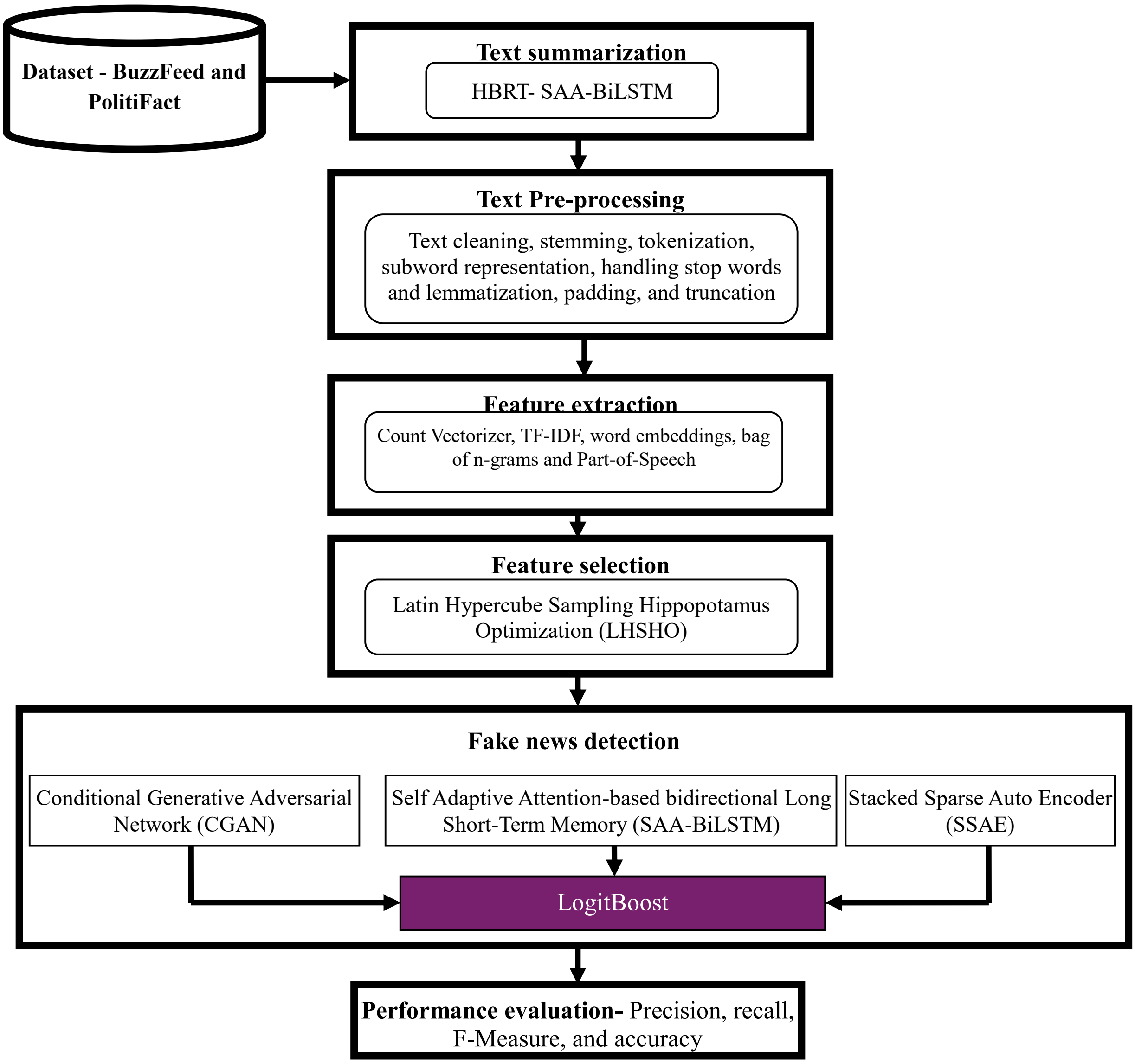


FIGURE 1. OVERALL PROCES OF PROPOSED SYSTEM

3.1. DATASET COLLECTION

Table 1, PolitiFact and BuzzFeed are used for experimentation. Totally 80% of samples are used for training, and 20% of samples are used for testing the model.

TABLE 1. DATASET DETAILS

News source	News articles	Fake news articles	Number of users
BuzzFeed	182	91	15257
PolitiFact	240	120	23865

3.2. TEXT SUMMARIZATION USING HBERT AND PRE-PROCESSING

HBERT is introduced to jointly capture fine-grained textual details and label semantics. In this approach, the pre-trained BERT model generates context-sensitive representations of fake news. To enhance this, an adaptive attention mechanism is applied, which learns label-specific word representations and identifies semantic relationships between words and labels. Finally, a hybrid representation merging both context-aware features and label-oriented word features is fed into a fake news encoder for effective detection. There are four main components of HBERT,

Word Embedding Module: Each word in the document is transformed into a context-sensitive representation through the embedding process.

Label Graph Embedding Module: The label graph serves as input for generating label embeddings, which capture and represent the semantic relationships among labels.

Adaptive Attention Module: This component computes attention scores between labels and words, thereby producing word representations that are specific to each label.

Aggregation Layer: At this stage, a hybrid representation is formed by combining context-aware features with label-specific word features, which is then used to summarize and integrate the most relevant information

Preprocessing is introduced to clean and eliminate unnecessary information from the dataset. It is the initial stage prior to training and text evaluation. Preprocessing aims to eliminate some of the irrelevant characteristics that set one dataset apart from another. Some text passages are not helpful for classifying the text as authentic or fraudulent. Preprocessing procedures are performed for the provided fake news dataset,

Text Cleaning: This step is considered as initial step of text preprocessing. The preprocessing step eliminates elements such as URLs, special symbols, numbers, emoticons, emojis, user mentions, and hashtags from the original tweet text.

Stemming: Stemming is utilized to reduce the number of words based on word type and class.

Tokenization and Subword Representation: BERT's tokenization method divides out-of-vocabulary words into smaller subunits, with special tokens indicating the start and finish of sequences.

Dealing with Stop Words and Lemmatization: Common stop words are eliminated, and lemmatization is applied to bring words back to their base forms while maintaining their meaning.

Padding and Truncation: To comply with HBERT's input size requirements, tweets are padded, while those exceeding the model's 512-token limit are truncated.

3.3. FEATURE EXTRACTION AND MATRIX CONSTRUCTION

Once the text has been preprocessed, it is necessary to extract features such as Count Vectorizer (CV), Term Frequency-Inverse Document Frequency (TF-IDF), word embeddings, bag of n-grams, and Part-of-Speech (POS) tagging.

Count Vectorizer (CV): CV, which is also referred to as term frequency, determines the token's rate value. It logs the frequency of each unique token in the text; the higher the token's score, the more often it occurs in the text.

TF-IDF: TF-IDF is a commonly used method to identify how important a word is within a document compared to a collection of documents. It works by combining two measures: term frequency (TF), which counts how often a word appears in a document, and inverse document frequency (IDF), which reduces the weight of words that appear too frequently across many documents. The final TF-IDF score represents each document as a vector in a high-dimensional space, giving more weight to words that are more meaningful in that context. Because of its ability to highlight significant terms, TF-IDF is widely used for feature extraction in text detection tasks

Word embeddings: The model starts with a word embedding module that maps input words into low-dimensional vectors. To generate context-aware representations, it employs the pre-trained HBERT model, which is built on a multi-layer bidirectional Transformer. This architecture allows the same word to be represented differently depending on its context. HBERT processes input sequences of up to 512 tokens and produces their corresponding representations. These learned embeddings are then used to represent words for detection model.

Bag of n-grams: Using the bag of n-grams method, a text is divided into its constituent words and then represented as a frequency distribution of those terms. Each sentence is transformed into overlapping N-grams by running a predetermined window of size N to create a vector utilizing N-grams.

PoS tagging: Assigning each word in a sentence to its proper grammatical category—noun, verb, adjective, and so on is known as Part-of-Speech (PoS) tagging, and it is an essential procedure. Understanding the structure of a sentence and identifying potential grammatical errors depend on knowing this syntactic information. PoS tagging is especially helpful for discovering syntactic infractions. Sentences with syntactical errors, such as misplaced modifiers or errors in subject-verb agreement can be identified by PoS tagging. By looking at the order of POS tags, models may identify patterns that differ from recognized grammatical structures, providing useful information for classifying sentences as acceptable or unacceptable.

In the News-User Engagement Matrix, user responses to specific news articles are illustrated. U represents the matrix, which has dimensions of $(n \times u)$, where n denotes the total number of news and u indicates the number of users. This matrix depicts the quantity of news stories shared by a specific person on social media. A heuristic approach for detecting communities in large-scale networks is introduced, which operates through an agglomerative, multi-step procedure. Community detection is studied in the context of small networks. This study created a user-community matrix based on available data and the user-user relationship. The process of matricization converts a

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2138-2152

tensor into a matrix by re-ordering it. Coupled Matrix–Tensor Factorization (CMTF) method, both news content and social context are represented together in a fused manner.

3.4. FEATURE SELECTION USING LATIN HYPERCUBE SAMPLING HIPPOPOTAMUS OPTIMIZATION (LHSHO)

Hippopotamus Optimization (HO) is inspired by three prominent behavioral patterns in the life of hippopotamuses, which are used for optimal selection of features from the dataset [36]. Hippopotamuses exhibit a defensive behavioral pattern that is activated when they are attacked by predators or when other creatures invade their territory, with optimal feature selection from OSN. Hippopotamuses show a defensive reaction by turning toward the predator, showcasing their formidable jaws while optimally selecting features from OSN. Carnivores like lions and spotted hyenas understand this phenomenon and take steps to evade direct exposure to potential harm when choosing the best attributes from OSN.

Population Initialization: Equation (1) is utilized to create a feature vector using OSN after the initial feature solution of HO is created at random,

$$x_{ij} = lb_j + rd \times (ub_j - lb_j), i = 1:N, j = 1:m \quad (1)$$

where $rd \in (0,1)$is a random number, x_{ij} is the position of i in the chosen feature solution, lb_j and ub_j are the limits of the j feature are indicated by equation (2),

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_N \end{bmatrix}_{N \times m} \quad (2)$$

where N is represented as the number of hippopotamuses in the herd, m is represented as count of features in the dataset.

Location Update (Exploration Phase): The dominant male will drive away male hippopotamuses. It will keep competing with other male hippopotamuses. The position of a male hippopotamus is indicated by Equation (3) [36],

$$x_{ij}^{M} = x_{ij} + y_1(D - I_1 x_{ij}), i = 1,2, \dots, \left[\frac{N}{2}\right], j = 1:m \quad (3)$$

Equation (5), x_{ij}^{M} is denoted as the location of the dominating hippopotamus is indicated by D, the location of the male hippopotamus is indicated by M, and $y_1 \in [0,1]$is random. In $I_1, I_2 \in [1,2]$ is denoted by equations (4)-(7),

$$h = \begin{cases} I_2 \times \overrightarrow{rd}_1 + (\sim Q_1) \\ 2 \times \overrightarrow{rd}_2 - 1 \\ \overrightarrow{rd}_3 \\ I_1 \times \overrightarrow{rd}_4 + (\sim Q_2) \\ rd_5 \end{cases} \quad (4)$$

$$T = \exp\left(-\frac{t}{Max_{iter}}\right) \quad (5)$$

$$x_{ij}^{FB} = \begin{cases} x_{ij} + h_1(D - I_2 MG_i)T > 0.6 \\ E, else \end{cases} \quad (6)$$

$$E = \begin{cases} x_{ij} + h_2(MG_i - D), rd_6 > 0.5 \\ lb_j + rd_7(ub_j - lb_j), else \end{cases} \quad (7)$$

where $T \in [0,1]$ is a random number, and $\overrightarrow{rd}_{1-4}, rd_{5-7}, \overrightarrow{rd}_{8-9} \in [0,1]$ is the random vector (Equation (5)). The ideal location of females for an ideal subset of features in the population is described by equations (5)–(6). According to equation (7), the young hippopotamus h is still close to the population if rd_6 is greater than 0.5; if not, it has already moved far away. The numbers or vectors h_1 and h_1 are chosen at random from equation (4),

$$x_i = \begin{cases} x_i^{M}, F_i^{M} < F_i \\ x_i, else \end{cases} \quad (8)$$

$$x_i = \begin{cases} x_i^{FB}, F_i^{FB} < F_i \\ x_i, else \end{cases} \quad (9)$$

The updated location of M within the population along with of extra weak hippopotamuses (words for fake news detection) are represented by Equations (8)-(9). F_i is denoted as the detection accuracy of classifier.

Hippo Defense against Predators: Hippopotamuses that are vulnerable within the population may exit the group in search of optimal features and become easy targets for both actual and fictitious news detection aimed at attack [36],

$$P_j = lb_j + \overrightarrow{rd}_8(ub_j - lb_j) \quad (10)$$

$$D = |P_j - x_{ij}| \quad (11)$$

Equation (11) is described as the distance from the killer to the i^{th} leader. The defensive behavior of hippos is centered on the F_{P_j} as a means of safeguarding themselves from predators while optimally selecting features from the dataset. If F_{P_j} is lower than F_i, it suggests that the danger of predation for hippopotamuses will rapidly shift toward the predator and compel them to withdraw. Should F_{P_j} be greater, equation (12) indicates that the predator's territory is more distant. It serves to alert predators not to trespass in their territory,

MSW MANAGEMENT -**Multidisciplinary, Scientific Work and Management Journal**
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2138-2152

$$x_{ij}^{HR} = \begin{cases} \overrightarrow{RL} \oplus P_j + \left(\frac{f}{\delta - d\cos(2\pi y)}\right) \cdot \left(\frac{1}{D}\right), F_{P_j} < F_i, i = \left[\frac{N}{2}\right] + 1, \dots, N, j = 1, \dots, m \\ \overrightarrow{RL} \oplus P_j + \left(\frac{f}{\delta - d\cos(2\pi y)}\right) \cdot \left(\frac{1}{2 \times D + \overrightarrow{rd}_9}\right), F_{P_j} \geq F_i \end{cases} \quad (12)$$

where x_{ij}^{HR} will encountering predators, a hippos position is known as HR. When a predator attacks a hippo, its stance is known as $\overrightarrow{RL}$. Equation (13), $\omega, v \in [0,1]$ are random numbers. Equation (14) is used to calculate σ_ω with the Levy distribution,

$$L(v) = 0.05 \times \frac{\omega \times \sigma_\omega}{|v|^{\frac{1}{v}}} \quad (13)$$

$$\sigma_\omega = \left[\frac{\Gamma(1+v)\sin\left(\frac{\pi v}{2}\right)}{\Gamma\left(\frac{1+v}{2}\right) v \times 2^{\frac{(v-1)}{2}}}\right]^{\frac{1}{v}} \quad (14)$$

Equation (12), $f \in [2,4], d \in [1,1.5], D \in [2,3], \delta \in [-1,1]$ is defined as the random numbers. Equation (15) states that the hippopotamus will be replaced in its optimal feature position if F_i^{HR} is greater than F_i; if not, it will re-enter the population.

$$x_i = \begin{cases} x_i^{HR}, F_j^{HR} < F_i \\ x_i, F_i^{HR} \geq F_i \end{cases} \quad (15)$$

Hippo Escape: Hippopotamuses also employ protective maneuvers to ward off predators when they encounter them creating a random position close to the hippopotamus present location. This behavior is being modeled using Equations (16)–(17). The number of iterations is denoted as t, maximum number of iterations is denoted as T. The hippopotamus shifts to a nearby safer region and updates its position if the new feature location results in a lower F_i [37],

$$ib_j^{local} = \frac{lb_j}{t}, ub_j^{local} = \frac{ub_j}{t}, t = 1,2, \dots \dots, T \quad (16)$$

$$x_{ij}^{HE} = x_{ij} + rd_{10}\left(lb_j^{local} + s_1\left(ub_j^{local} - lb_j^{local}\right)\right) \quad (17)$$

In (17), x_{ij}^{HE} is worked with searching for the optimal feature position of the h to discover the nearest protected spot. Equation (18) demonstrates enhanced local search capabilities as s_1,

$$s_1 = \begin{cases} 2 \times \overrightarrow{rd}_{11} - 1 \\ rd_{12} \\ rd_{13} \end{cases} \quad (18)$$

In Equation (18), $\overrightarrow{rd}_{11}, rd_{12-13}$ is represented as a random vector [0, 1] which stay to a normal distribution,

$$x_i = \begin{cases} x_i^{HE}, F_i^{HE} < F_i \\ x_i, F_i^{HE} \geq F_i \end{cases} \quad (19)$$

In HO, for the purpose of optimal feature selection, Latin hypercube sampling (LHS) is used in place of random initialization during the generation of the initial population. This result in a more uniform initial population compared to that produced by random numbers. LHS is a technique for uniform sampling within a feature selection search space, aimed at producing a set of features that are as uniform and infrequently repeated as possible across each parameter space. LHS process is described as follows,

(1) Select a point at random from each interval after evenly dividing [0, 1] into n.

(2) To guarantee that the news articles on every parameter axis are uniformly circulated and not duplicated, the positions of each news articles are alternated at random. Equation (20) demonstrates how the Latin Hypercube Optimization (LHO) method generates a uniform distribution over the interval (0,1), which is then utilized for initializing the population in the news article dataset,

$$Cootpos(i) = (ub - lb) * lhs(1, D) + lb \quad (20)$$

LHS is used for population initialization helps prevent the population from becoming trapped at this stage. Local optimal solutions enhance the ability of algorithm to conduct global searches. Additionally, the broader initial distribution of the population aids the algorithm in exploring enhanced detection accuracy and improving its efficiency in locating the optimal solution. The overall flow process of LHSHO is depicted in Figure 2.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2138-2152

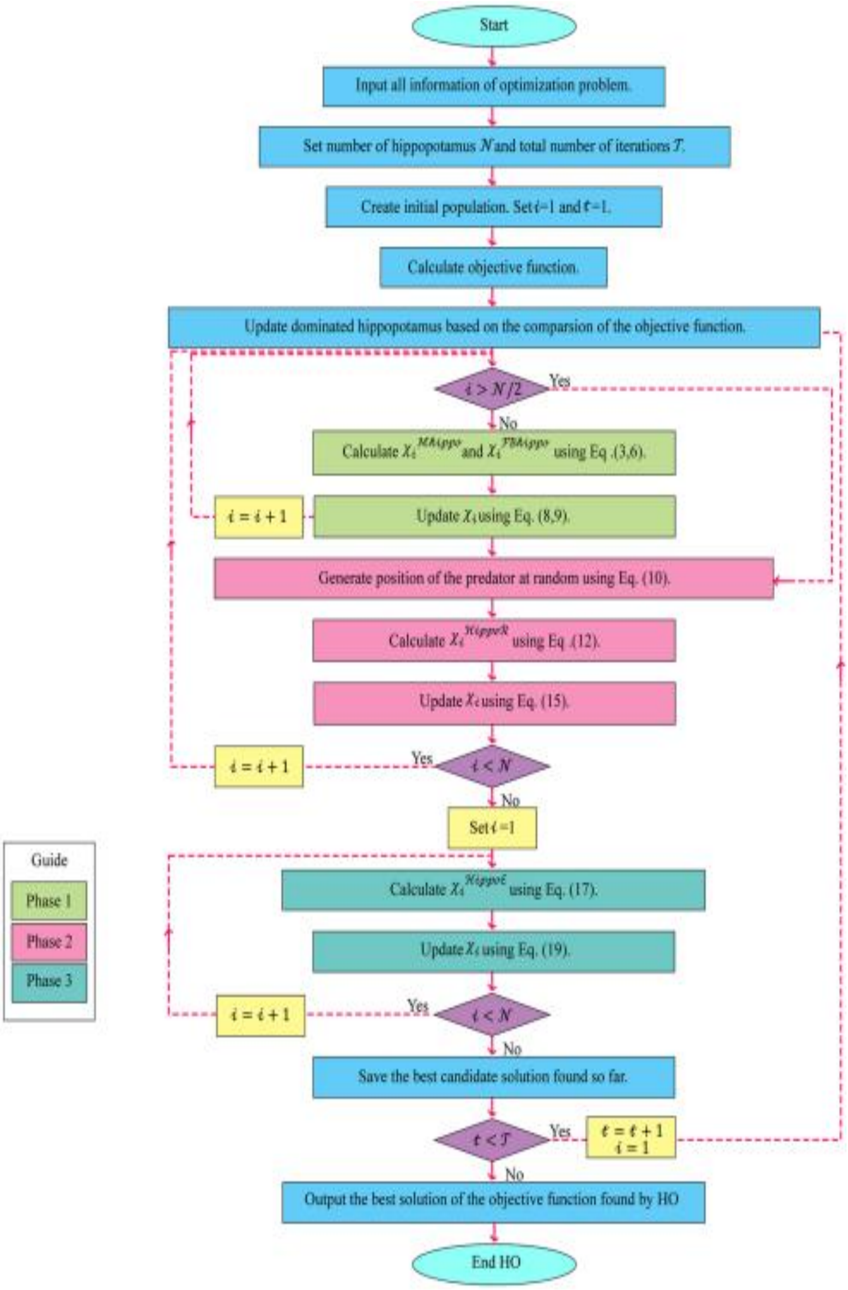


FIGURE 2. FLOWCHART OF LATIN HYPERCUBE SAMPLING HIPPOPOTAMUS OPTIMIZATION (LHSHO)

ALGORITHM 1. LATIN HYPERCUBE SAMPLING HIPPOPOTAMUS OPTIMIZATION (LHSHO)

INPUT: Population (feature) size (N), Maximum number of iteration (Max_{iter}), Number of dimension

OUTPUT: Optimal subset of features

Start procedure

1. Initialize the parameters based equation (1)
2. Calculate of Fitness value(classification accuracy)
3. For $t = 1{:}T$

Phase 1: hippopotamuses position update in the river

4. For $i = 1{:}N/2$
5. Calculate the new feature position for ith **hippopotamuses** using equations (3) (6) (Proposed method)
6. Update feature position of ith **hippopotamuses** using equations (8)-(9)
7. End For

Phase 2: Hippopotamus defence against predators

8. For $i = 1 + N/2{:}N$
9. Generate random position for predator using equation (10)

10. Calculate the new position for i[th] hippopotamuses using equation (12)
11. Calculate the position for i[th] hippopotamuses using equation (15)
12. End For

Phase 3: Hippopotamus Evading the Predator (Exploitation)

13. Calculate the new bound of features based on equation (16)
14. For $i = 1:N$
15. Generate random position of features using equation (17)
16. Update the new feature position of i[th] hippopotamuses using equation (19)
17. End For
18. Save the best feature solution found so far
19. End For
20. Output the optimal best feature solution of the objective function

End procedure

3.5. STACKING ENSEMBLE DEEP LEARNING (SEDL) MODEL

To enhance detection performance, stacking is an ensemble strategy that combines methods like as CGAN, SAA-BiLSTM, and SSAE. Logistic Regression (LR) is trained which learns to create detection depending on the detection of the single methods. The general flow of SEDL model is depicted in Figure 3.

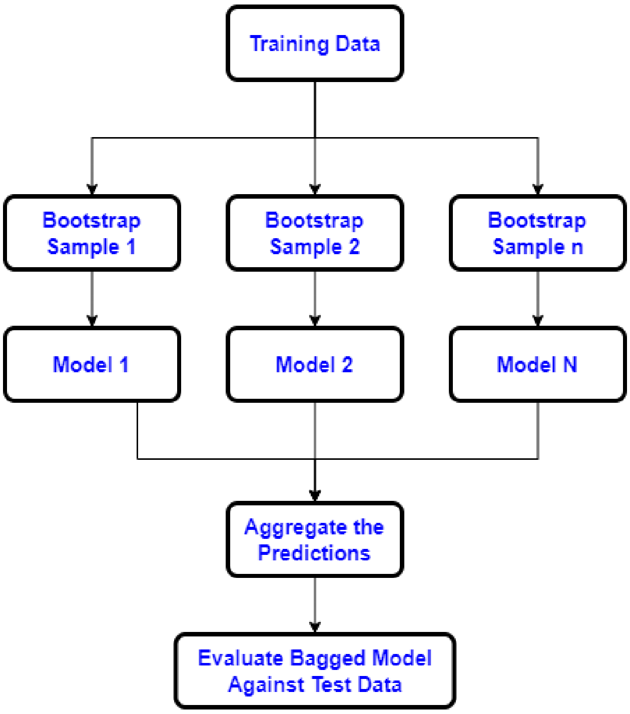


FIGURE 3. STACKING ENSEMBLE DEEP LEARNING (SEDL) MODEL

3.5.1. Conditional Generative Adversarial Network(CGAN)

Generative Adversarial Network (GAN) is introduced to generate real samples and fake news. The fundamental idea of the discriminator (D) and generator (G) networks compete with one another. To differentiate genuine samples from the database and artificial ones produced by G, the generator takes random inputs and tries to deceive the discriminator D. The data representation is mapped from a dZdZ-dimensional noise space Z into the data space X through $G{:}Z\rightarrow X$. Equation (21) presents the min–max objective function, where the discriminator D estimates the probability that a given input originates from the real data distribution, defined as $D{:}X\rightarrow[0,1]$,

$$\min_G \max_D V(D,G) = E_D + E_G \tag{21}$$

$$E_D = E_{x\sim p_{data}(x)}[log\ D(x)] \tag{22}$$

$$E_G = E_{z\sim p_z(z)}\left[log\left(1 - D\left(G(z)\right)\right)\right] \tag{23}$$

where x∈X and z∈Z are sampled from the data distribution $p_{data}(x)$ and the noise distribution $p_z(z)$ respectively. During training, the discriminator D and generator G are updated alternately in successive iterations [38]. The projection-based variant of a conditional adversarial network is referred to as CGAN [38], where an additional label space Y is introduced to incorporate class information from the training dataset.

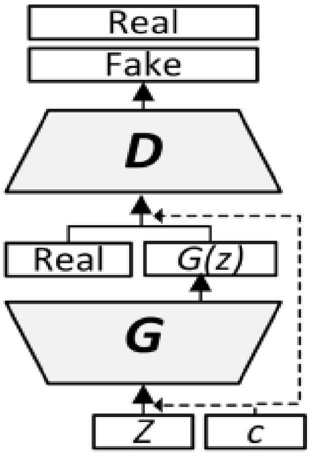


FIGURE 4. ARCHITECTURES OF THE CGAN

The equation (24) is a description of the modified D and G,

$$G: Z \times Y \to XD: X \times Y \to [0,1] \tag{24}$$

Equation (25) is used as the loss function [39],

$$\min_G \max_D V(D, G) = E_D + E_G \tag{25}$$

where,

$$\min_G \max_D V(D, G) = E_D + E_G \tag{26}$$

$$E_D = E_{x,y \sim p_{data}(x,y)}[\log D(x, y)] \tag{27}$$

$$E_G = E_{z \sim p_z(z),p(y)}\left[\log\left(1 - D(G(z, y), y)\right)\right] \tag{28}$$

3.5.2. SAA-BiLSTM

SAA-BiLSTM at order to prevent the loss of crucial information, an attention mechanism is added to determine the correlation between hidden features at each time step. The resemblance between two sets of real and fake news is determined using Self Adaptive Attention (SAA). The correlation is normalized to determine the weights of the values that correlate to the gates, and the values are then weighted to provide the final result. Together with previous quality data, the dataset first travels via the input layer before being sent to the SAA-BiLSTM cell structure. The individual gate structures in this cell start optimizing information that is critical to quality features in order to choose which information should be remembered and which should be forgotten. In situations when the sequence data is too lengthy, the SAA-BiSLTM may be able to identify the relationship between each time step and prevent the loss of crucial information pertaining to the present detection by providing the attention layer with crucial information in a disguised state. As anticipated by the existing input feature hidden state wholly linked weights and biases, the sequence hidden state enters the completely connected layer utilizing the weighted summation of attention weights. Furthermore, there is a clear correlation between the z value of the SAA-BiLSTM model and the dynamics and complexity of fake news detection.

3.5.3. Stacked Sparse Auto Encoder (SSAE)

An Autoencoder (AE) is a three-layer neural network consisting of an input layer, a hidden layer, and an output layer, and it is often applied in fake news detection. The AE progressively converts specific feature vectors into more abstract ones, enabling a nonlinear transformation from high-dimensional data to a lower-dimensional space. Sparse coding, originally introduced to mimic the receptive field learning of simple cells [40]. It is incorporated into the AE by adding a sparsity penalty to the hidden layer. This constraint helps extract more compact and efficient low-dimensional features, making the input data representation more meaningful. Let the average activation of neurons in the hidden layer be denoted as $\hat{\rho}_j$ in equation (29),

$$\hat{\rho}_j = \frac{1}{N}\sum_{i=1}^{N}[n_j(x_i)] \tag{29}$$

Error function of AE is regularized by adding Kullback Leibler (KL) divergence in equation (30),

$$KL(\rho \| \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \tag{30}$$

Overall objective function of SSAE is described by equation (31),

$$J_{sparse}(W, b) = J(W, b) + \mu \sum_{j=1}^{m} KL\left(\rho \| \hat{\rho}_j\right) \tag{31}$$

where μ is a weighting factor, and m is the total number of hidden units. Equation (32) weight attenuation is applied to the error function to minimize overfitting. The attenuation coefficient of the weight is denoted by λ,

$$J_{sparse}(W, b) = J_E(W, b) + \mu \sum_{J=1}^{M} KL\left(\rho \| \hat{\rho}_j\right) + \frac{\lambda}{2}\sum_{r=1}^{3}\sum_{i=1}^{m}\sum_{j=1}^{m+1}\left(w_{ij}^{r}\right)^2 \tag{32}$$

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2138-2152

Figure 5 illustrates the topology of a stacked sparse auto encoder (SSAE) neural network, which is made up of several SAE connected end to end. Higher-level feature representations of the samples are created by using the output of the previous layer of SAE as the input of the subsequent layer of the self-encoder. The optimal weights and biases of the SSAE are obtained step by step using a greedy layer-wise pre-training strategy. Then it is fine-tuned using the error back propagation approach until the lesser error is attained among input news and the output detection. Regarding the SAE-defined error function $J_{sparse}(W,b)$ by equations (33)-(34),

$$\frac{\partial}{\partial w_{ij}^{r}} J_{sparse}(W,b) = \frac{1}{2n_r}\sum_{r=1}^{n_r}\frac{\partial}{\partial w_{ij}^{r}} J_{sparse}(W,b,X(n),Y(n)) + \lambda w_{ij}^{r} \tag{33}$$

$$\frac{\partial}{\partial b^{r}} J_{sparse}(W,b) = \frac{1}{2n_r}\sum_{r=1}^{n_r}\frac{\partial}{\partial b^{r}} J_{sparse}\big(W,b,X(n),Y(n)\big) \tag{34}$$

Consequently, the weight and bias updating procedure is described by equations(35)-(36),

$$w_{ij}^{k} = w_{ij}^{k} - \eta\frac{\partial}{\partial w_{ij}^{k}} J(W,b) \tag{35}$$

$$b^{r} = b^{r} - \eta\frac{\partial}{\partial b^{r}} J(W,b) \tag{36}$$

where η is the update learning rate and X(n) and Y (n) are the n^{th} actual vector and its reconstruction vector, correspondingly.

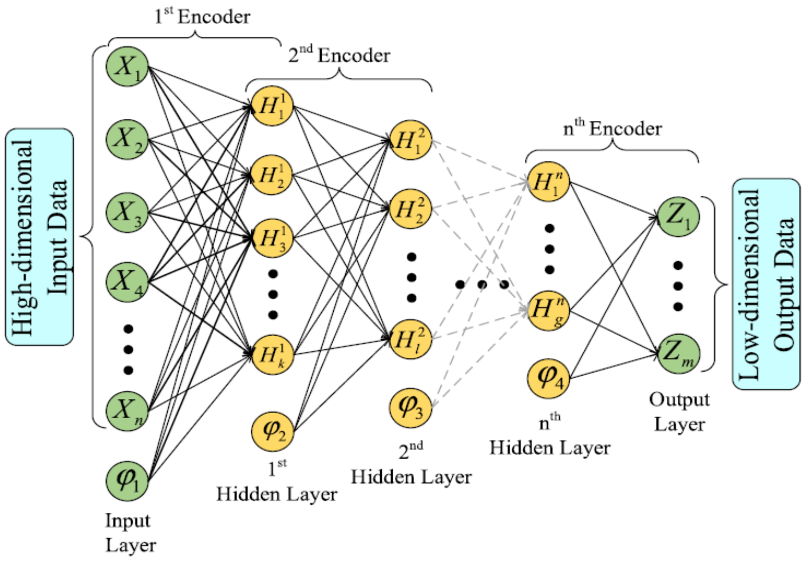


FIGURE 5. STRUCTURE OF STACKED SPARSE AUTO ENCODER (SSAE) MODEL

SSAE network, the dynamic adaptive modification of various parameters is accomplished by the application of adaptive moment estimation (Adam). Equations (37)–(39) illustrates the Adam algorithm which calculates the gradient first-order moment estimate m_t and second-order moment estimate v_t, their damping decrements are represented as β_1, β_2. The gradient of the parameters in the loss function at time step t is denoted by g_t,

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\cdot g_t \tag{37}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\cdot g_t^2 \tag{38}$$

$$g_t \leftarrow \nabla_\theta J_t(\theta_{t-1}) \tag{39}$$

Equations (40–41) is used for computer bias of $m_t$and v_t,

$$m_t' = \frac{m_t}{1-\beta_1^t} \tag{40}$$

$$v_t' = \frac{v_t}{1-\beta_2^t} \tag{41}$$

Parameters are updated using equation (42),

$$\theta_{t+1} = \theta_t - \frac{\gamma}{\sqrt{v_t'}+\varepsilon}\cdot m_t' \tag{42}$$

To keep the denominator from being zero, ε takes a little constant, and γ is the update stepsize.

3.5.4. Stacking Ensemble Classifier

Stacking is an ensemble learning strategy where a second level learner, sometimes known as a meta-learner, is used to consolidate predictions from a single method. Two levels, such as level-0 and level-1, make up the stacking structure, and the meta-learner (level-1) combines the outputs of several base learners (level-0). Assume that examples $d_i = (x_i, y_i)$, where $x_i$denotes conditioning factors and $y_i$denotes associated classifications, make up the initial dataset D. where N is the total number of news, and $i \in [1, N]$, In order to create

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2138-2152

level-0 classifiers, the dataset D is first repeatedly split into two disjoint subsets, one of which is used to train base learning algorithms designated as h_t,

$$h_t^i = L_t(D - d_i)\ \forall i = 1,2,\ldots,N, \forall t = 1,2,3 \quad (43)$$

Using trained classifiers, the remaining instances are utilized to generate predictions (z_{it}).

$$z_{it} = h_t^i(x_i) \quad (44)$$

The new dataset $D' = (z_{it}, y_i)$, is created by combining these outputs from level-0 classifiers with their actual classification. The meta-learner (Logistic Regression (LR)) is then trained using this dataset.

$$Y_x = LR(h_1(x), h_2(x), h_3(x)) \quad (45)$$

To generate the final detection, LR can compile the basis learners' classification results.

4. RESULTS AND DISCUSSION

In this section, experimentation is performed using MATrix LABoratory R2023a (MATLABR 2023a). BERT model has been developed using Python 3.11 with Google Colab. These methods are experimented with Intel Core i5/i7 processor, NVIDIA CUDA-enabled GPU, 8 GB RAM, Windows 10 OS. The results of various detection methods like EchoFakeD, DNN, CNN-BiLSTM, Bi-LSTM-GRU, and SAA-BiLSTM has been experimented using benchmark datasets (BuzzFeed and PolitiFact). Textual content of news articles, user-context, and user-based relationships is served as basis for these techniques. Precision, recall, f-measure, and accuracy are used to validate the effectiveness of the detection methods.

Precision: Precision measures the proportion of appropriately identified real news out of the total instances recognized as real.It is given by equation (46).

$$\text{Precision} = \frac{\text{True Positive (TP)}}{(\text{True Positive (TP)} + \text{False Positive (FP)})} \quad (46)$$

Recall: Recall is the fraction of properly predictable real cases among all actual real instances. It is given by equation (47),

$$\text{Recall} = \frac{\text{True Positive (TP)}}{(\text{True Positive (TP)} + \text{False Negative (FN)})} \quad (47)$$

F-Measure: F-Measure as the harmonic mean of precision and recall by equation (48),

$$F - \text{Measure} = \frac{2 * \text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} \quad (48)$$

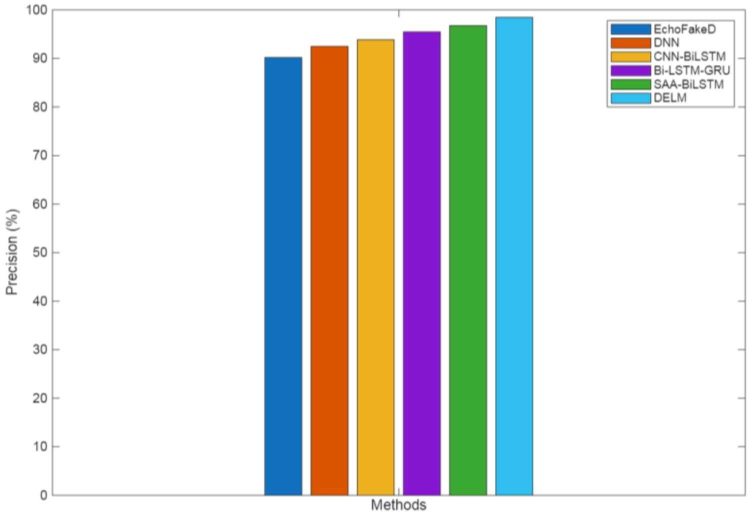
Accuracy: Equation (49), accuracy is the total of all correctly classified data, whether they were authentic or not,

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (49)$$

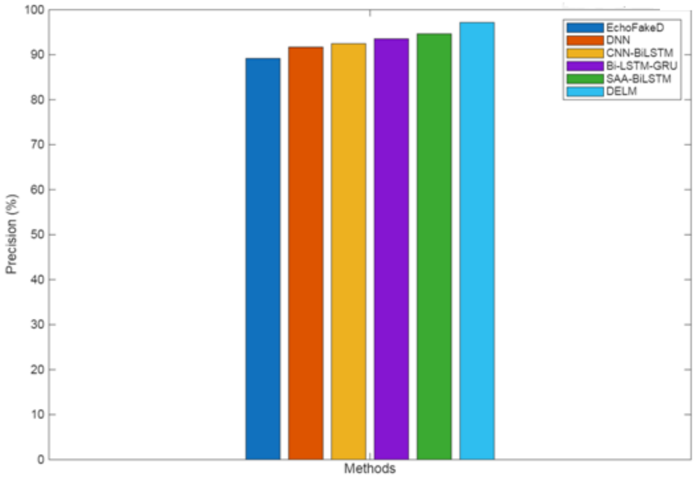
In this case, TP is denoted as the correctly identified instances TN is denoted as the accurately rejected instances, FP is denoted as the incorrectly identified instances, and FN is denoted as the incorrectly rejected instances. To verify the effectiveness of the detection techniques using assessment metrics like as precision, recall, f-measure, and accuracy. The detection results of BuzzFeed and PolitiFact datasets against classification methods are displayed in Table 2. The highest precision, recall, f-measure, and accuracy results for the DELM classifier on the Buzzfeed dataset are 98.45%, 98.00%, 98.23%, and 98.29%, respectively. The results of the classification show that the model and feature selection are crucial in improving the detection outcomes.

TABLE 2. EVALUATION ANALYSIS OF DETECTION METHODS (BUZZFEED AND POLITIFACT)

METHODS	BUZZFEED (%)			
	PRECISION	RECALL	F-MEASURE	ACCURACY
EchoFakeD	90.20	92.06	91.12	91.49
DNN	92.48	94.15	93.31	93.62
CNN-BiLSTM	93.85	95.08	94.46	94.83
Bi-LSTM-GRU	95.47	96.33	95.90	96.07
SAA-BiLSTM	96.76	97.35	97.05	97.16
DELM	98.45	98.00	98.23	98.29
METHODS	POLITIFACT (%)			
	PRECISION	RECALL	F-MEASURE	ACCURACY
EchoFakeD	89.18	92.42	90.77	91.76
DNN	91.69	93.08	92.38	93.53
CNN-BiLSTM	92.47	94.10	93.27	94.61
Bi-LSTM-GRU	93.54	95.38	94.45	95.56
SAA-BiLSTM	94.65	97.50	96.06	96.47
DELM	97.17	98.75	97.95	98.23



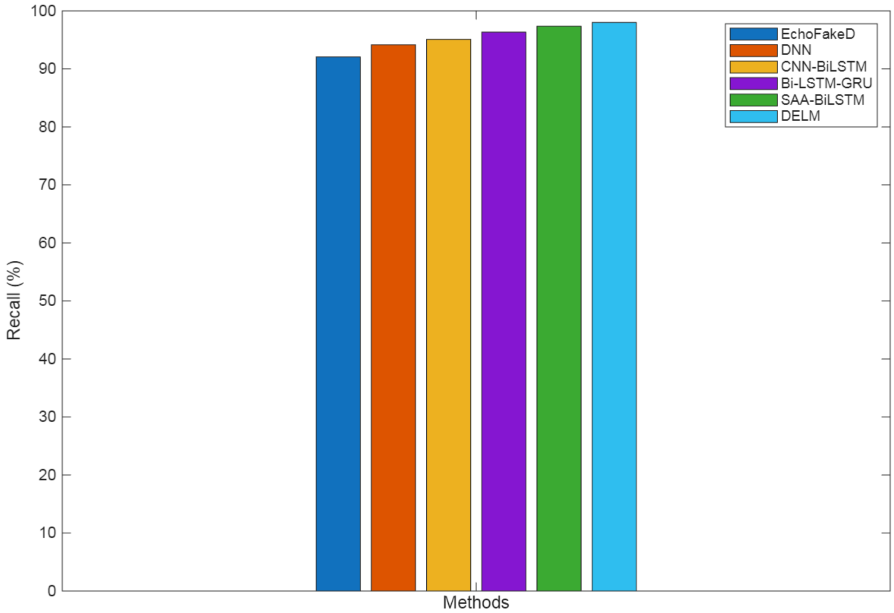
(a) Precision results of Buzzfeed dataset



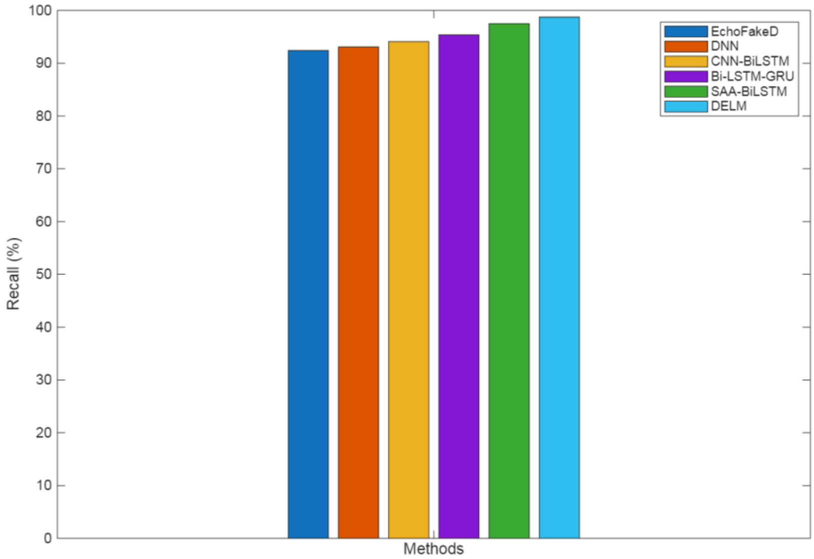
(b) Precision results of Politifact dataset

FIGURE 6. PRECISION COMPARISON AGAINST METHODS (FAKE NEWS DATASETS)

Figure 6(a)&(b), precision results are evaluated using existing methods, and proposed method. DELEM has highest results of 98.45%, and 97.17% for Buzzfeed and Politifact. Buzzfeed, EchoFakeD, DNN, CNN-BiLSTM, Bi-LSTM-GRU, and SAA-BiLSTM provide lowest precision of 90.20%, 92.48%, 93.85%, 95.47%, and 96.76% are illustrated in Figure 6(a). Politifact, EchoFakeD, DNN, CNN-BiLSTM, Bi-LSTM-GRU, and SAA-BiLSTM provide lowest precision of 89.18%, 91.69%, 92.47%, 93.54% and 94.65% are illustrated in Figure 6(b). Irrelevant features of samples are reduced using LHSHO algorithm. DELM algorithm increases the precision results by combining the results of CGAN, SAA-BiLSTM, and SSAE.



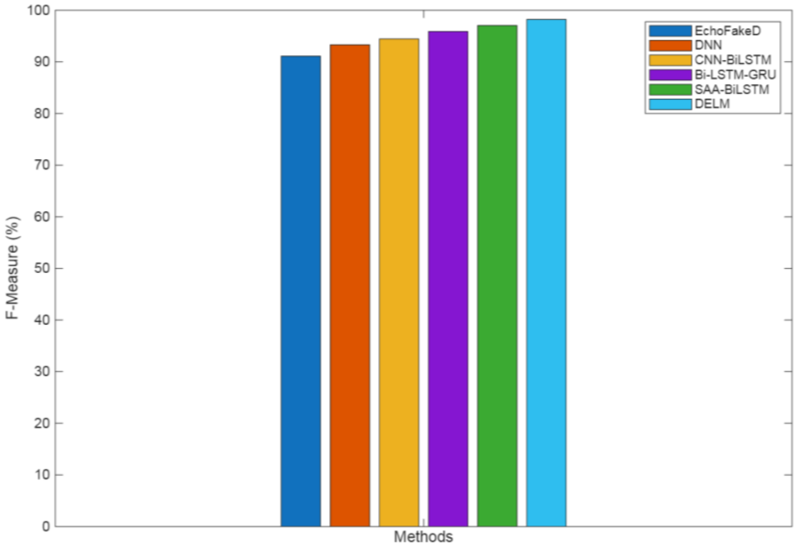
(a) Recall results of Buzzfeed dataset



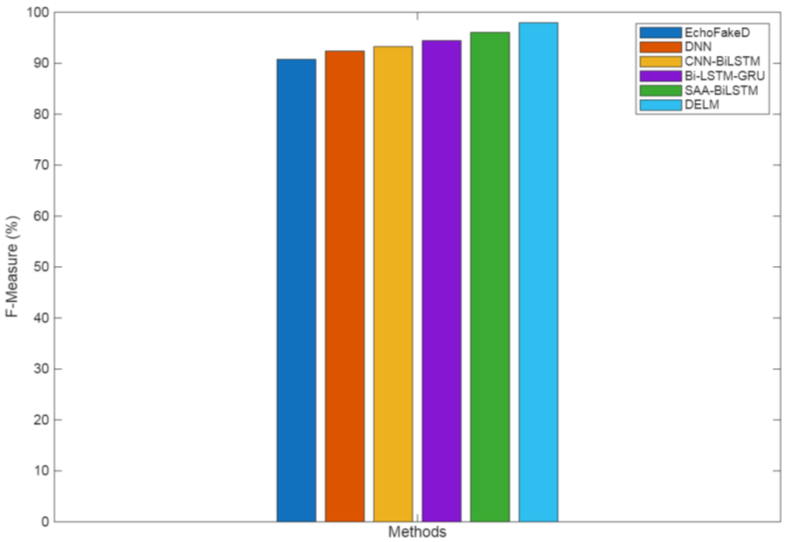
(b) Recall results of Politifact dataset

FIGURE 7. RECALL COMPARISON AGAINST METHODS (FAKE NEWS DATASETS)

Recall analysis is compared between the detection methods are illustrated in Figure 7(a)&(b). DELEM has highest results of 98.00%, and 98.75% for Buzzfeed and Politifact. Buzzfeed, EchoFakeD, DNN, CNN-BiLSTM, Bi-LSTM-GRU, and SAA-BiLSTM provide lowest recall of 92.06%, 94.15%, 95.08%, 96.33%, and 97.35% are illustrated in Figure 7(a). Politifact, EchoFakeD, DNN, CNN-BiLSTM, Bi-LSTM-GRU, and SAA-BiLSTM provide lowest recall of 92.42%, 93.08%, 94.10%, 95.38% and 97.50% are illustrated in Figure 7(b). DELM algorithm increases the recall results since it combines the procedure of several classification methods.



(a) F-Measure results of Buzzfeed dataset

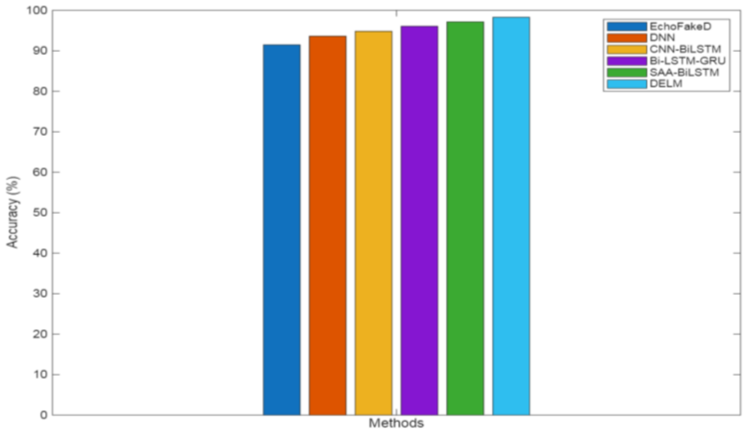


(b) F-Measure results of Politifact dataset

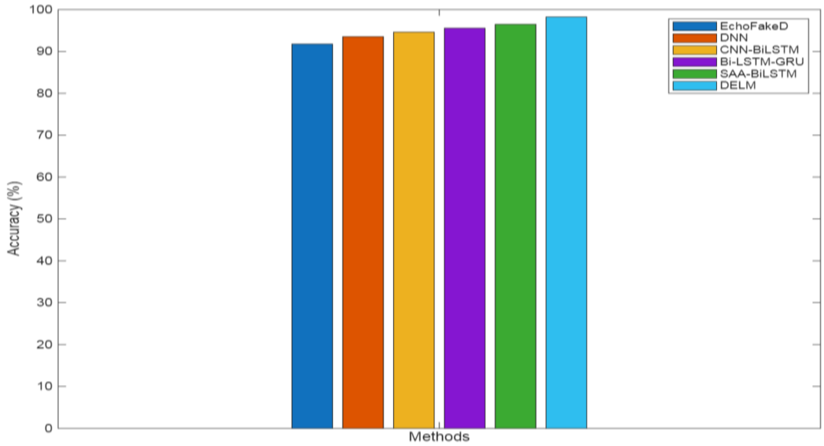
MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages: 2138-2152

FIGURE 8.F-MEASURE COMPARISON AGAINST METHODS (FAKE NEWS DATASETS)

Proposed model and existing methods in terms of f-measure are illustrated in figure 8(a)&(b). DELEM has highest results of 98.23%, and 97.95% for Buzzfeed and Politifact. Buzzfeed, EchoFakeD, DNN, CNN-BiLSTM, Bi-LSTM-GRU, and SAA-BiLSTM provide lowest f-measure of 91.12%, 93.31%, 94.46%, 95.90%, and 97.05% are illustrated in Figure 8(a). Politifact, EchoFakeD, DNN, CNN-BiLSTM, Bi-LSTM-GRU, and SAA-BiLSTM provide lowest f-measure of 90.77%, 92.38%, 93.27%, 94.45% and 96.06% are illustrated in Figure 8(b).



(a) Accuracy results of Buzzfeed dataset



(b) Accuracy results of Politifact dataset

FIGURE 9.ACCURACY COMPARISON AGAINST METHODS (FAKE NEWS DATASETS)

Proposed model and existing methods in terms of accuracy are illustrated in figure 9(a)&(b). EchoFakeD, DNN, CNN-BiLSTM, Bi-LSTM-GRU and SAA-BiLSTM gives lowest accuracy of 91.49%, 93.62%, 94.83%, 96.07% and 97.16% for BuzzFeed are illustrated in Figure 9(a). EchoFakeD, DNN, CNN-BiLSTM, Bi-LSTM-GRU and SAA-BiLSTM gives lowest accuracy of 91.76%, 93.53%, 94.61%, 95.56% and 96.47% for PolitiFact are illustrated in Figure 9(b). DELM algorithm provides higher accuracy of 98.29% and 98.23% for BuzzFeed and PolitiFact datasets.

5. CONCLUSION AND FUTURE WORK

In this work, novel feature selection and classification model is introduced for fake news detection. Hybrid Bidirectional Encoder Representation from Transformers (HBERT) model is introduced to compute context-aware representation for text summarization. Then, text preprocessing and feature extraction methods are introduced for summarized text. Feature selection is performed using Latin Hypercube Sampling Hippopotamus Optimization (LHSHO) algorithm encompasses the hippopotamus' instinctual response of fleeing from predators and dynamically in search of distance itself from areas of possible danger with best selection of features from OSN. LHS is a uniform sampling technique which is applied to generate feature sets that are evenly distributed and minimally repetitive within each parameter dimension. By ensuring a well-distributed initial population, the algorithm improves better exploration capability, which enhances detection accuracy and improves efficiency by reducing the chances of being trapped in local optima. Finally, Stacking Ensemble Deep Learning (SEDL) model is introduced by utilizing Conditional Generative Adversarial Network (CGAN), Self Adaptive Attention-based bidirectional Long Short-Term Memory (SAA-BiLSTM), and Stacked Sparse Auto Encoder (SSAE) for the fake news detection. Stacking is used to leverage the strengths of these methods by training a Logistic Regression (LR) meta-model. It is used to make decision depending on the results of the individual classifiers. BuzzFeed and PolitiFact datasets has been used to evaluate methods using precision, recall, F-Measure, and accuracy metrics. In the future, present system is implemented with high speed and low complexity against OSN. The present model has been also implemented to huge datasets, real-time OSNs with more number of labels, special characters, numeric values and emoticons in the news articles.

REFERENCES

1. Keijzer, M.A., Mäs, M. and Flache, A., 2018. Communication in online social networks fosters cultural isolation. Complexity, 2018(1), pp.1-18.
2. Ciampaglia, G. L. Fighting fake news: A role for computational social science in the fght against digital misinformation. J. Comput. Soc. Sci. 1, 147–153.
3. Gravanis, G., Vakali, A., Diamantaras, K. and Karadais, P., 2019. Behind the cues: A benchmarking study for fake news detection. Expert Systems with Applications, 128, pp.201-213.
4. Faustini, P.H.A. and Covoes, T.F., 2020. Fake news detection in multiple platforms and languages. Expert Systems with Applications, 158, pp.1-27.

5. Shrivastava, G.; Kumar, P.; Ojha, R.P.; Srivastava, P.K.; Mohan, S.; Srivastava, G. Defensive modeling of fake news through online social networks. IEEE Trans. Comput. Soc. Syst. **2020**, 7, 1159–1167.
6. Xu, K.; Wang, F.; Wang, H.; Yang, B. Detecting fake news over online social media via domain reputations and content understanding. Tsinghua Sci. Technol. **2019**, 25, 20–27.
7. Habib, A., Asghar, M.Z., Khan, A., Habib, A. and Khan, A., 2019. False information detection in online content and its role in decision making: a systematic literature review. Social Network Analysis and Mining, 9(1), pp.1-20.
8. Xarhoulacos, C.G., Anagnostopoulou, A., Stergiopoulos, G. and Gritzalis, D., 2021. Misinformation vs. situational awareness: The art of deception and the need for cross-domain detection. Sensors, 21(16), pp.1-41.
9. Ciampaglia, G.L., 2018. Fighting fake news: a role for computational social science in the fight against digital misinformation. Journal of Computational Social Science, 1(1), pp.147-153.
10. Guo, M., Xu, Z., Liu, L., Guo, M. and Zhang, Y., 2020. An adaptive deep transfer learning model for rumor detection without sufficient identified rumors. Mathematical Problems in Engineering, 2020(1), pp.1-12.
11. Varshney, D.; Vishwakarma, D.K. Vishwakarma, Hoax news-inspector: A real-time prediction of fake news using content resemblance over web search results for authenticating the credibility of news articles. J. Ambient Intell. Humaniz. Comput. 2020, 12, 8961–8974.
12. Ghosh S, Shah C (2018) Towards automatic fake news classification. Proc Assoc Inf Sci Technol 55(1):805–807.
13. Sharma K, Qian F, Jiang H, Ruchansky N, Zhang M, Liu Y (2019) Combating fake news: a survey on identification and mitigation techniques. ACM Trans Intell Syst Technol 10(3):1–42.
14. Shu K, Mahudeswaran D, Wang S, Lee D, Liu H (2020) Fake- NewsNet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data 8(3):171–188.
15. Dilawari A., M.U.G. Khan, ASoVS: abstractive summarization of video sequences, IEEE Access 7 (2019) 29253–29263.
16. Patil S.P., L. Chavan, J. Mukane, D.R. Vora, V. Chitre, State-of-the-art approach to e-learning with cutting edge NLP transformers: Implementing text summarization, question and distractor generation, question answering, Int. J. Adv. Comput. Sci. Appl. (2022).
17. Su, M.H., Wu, C.H. and Cheng, H.T., 2020. A two-stage transformer-based approach for variable-length abstractive summarization. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, pp.2061-2072.
18. Chen, Y., Ma, Y., Mao, X. and Li, Q., 2019. Multi-task learning for abstractive and extractive summarization. Data Science and Engineering, 4(1), pp.14-23.
19. Qaroush A., I.A. Farha, W. Ghanem, M. Washaha, E. Maali, An efficient single document Arabic text summarization using a combination of statistical and semantic features, J. King Saud Univ.-Comp. Inform. Sci. 33 (6) (2021) 677–692.
20. Areshey, A. and Mathkour, H., 2023. Transfer learning for sentiment classification using bidirectional encoder representations from transformers (BERT) model. Sensors, 23(11), pp.1-18.
21. Gorenstein, L., Konen, E., Green, M. and Klang, E., 2024. Bidirectional encoder representations from transformers in radiology: a systematic review of natural language processing applications. Journal of the American College of Radiology, 21(6), pp.914-941.
22. Yang, Y., Lv, H. and Chen, N., 2023. A survey on ensemble learning under the era of deep learning. Artificial Intelligence Review, 56(6), pp.5545-5589.
23. Shaikh, T.A., Rasool, T., Verma, P. and Mir, W.A., 2024. A fundamental overview of ensemble deep learning models and applications: systematic literature and state of the art. Annals of Operations Research, pp.1-77.
24. Szczepański, M., Pawlicki, M., Kozik, R. and Choraś, M., 2021. New explainability method for BERT-based model in fake news detection. Scientific reports, 11(1), pp.1-13.
25. Elsaeed, E., Ouda, O., Elmogy, M.M., Atwan, A. and El-Daydamony, E., 2021. Detecting fake news in social media using voting classifier. IEEE Access, 9, pp.161909-161925.
26. Jain, M.K., Gopalani, D. and Meena, Y.K., 2025. Hybrid CNN-BiLSTM model with HHO feature selection for enhanced fake news detection. Social Network Analysis and Mining, 15(1), p.43.
27. Verma, A., Priyanka, P., Khan, T., Singh, K., Yesufu, L.O., Ariffin, M.M. and Ahmadian, A., 2025. ScrutNet: a deep ensemble network for detecting fake news in online text. Social Network Analysis and Mining, 15(1), pp.1-16.
28. Aslam, N., Ullah Khan, I., Alotaibi, F.S., Aldaej, L.A. and Aldubaikil, A.K., 2021. Fake detect: A deep learning ensemble model for fake news detection. complexity, 2021(1), pp.1-8.
29. Venkatachalam, K., Al-Onazi, B.B., Simic, V., Tirkolaee, E.B. and Jana, C., 2023. DeepFND: an ensemble-based deep learning approach for the optimization and improvement of fake news detection in digital platform. PeerJ Computer Science, 9, pp.1-25.
30. Qu, Z., Meng, Y., Muhammad, G. and Tiwari, P., 2024. QMFND: A quantum multimodal fusion-based fake news detection model for social media. Information Fusion, 104, p.102172.
31. Kaliyar, R.K., Goswami, A. and Narang, P., 2019, Multiclass fake news detection using ensemble machine learning. In 2019 IEEE 9th international conference on advanced computing (IACC), pp. 103-107.
32. Ali, A.M., Ghaleb, F.A., Al-Rimy, B.A.S., Alsolami, F.J. and Khan, A.I., 2022. Deep ensemble fake news detection model using sequential deep learning technique. Sensors, 22(18), pp.1-18.
33. Padalko, H., Chomko, V., Yakovlev, S. and Chumachenko, D., 2023. Ensemble machine learning approaches for fake news classification. Radioelectronic and Computer Systems, (4), pp.5-19.
34. Jiang, T.A.O., Li, J.P., Haq, A.U., Saboor, A. and Ali, A., 2021. A novel stacking approach for accurate detection of fake news. IEEE Access, 9, pp.22626-22639.
35. Kaliyar, R.K., Goswami, A., Narang, P. and Sinha, S., 2020. FNDNet–a deep convolutional neural network for fake news detection. Cognitive Systems Research, 61, pp.32-44.
36. Wang, H., Binti Mansor, N.N. and Mokhlis, H.B., 2024. Novel hybrid optimization technique for solar photovoltaic output prediction using Improved Hippopotamus Algorithm. Applied Sciences, 14(17), pp.1-50.
37. Han, T., Wang, H., Li, T., Liu, Q. and Huang, Y., 2025. MHO: A Modified Hippopotamus Optimization Algorithm for Global Optimization and Engineering Design Problems. Biomimetics, 10(2), pp.1-31.
38. Wang, X., Ghasedi Dizaji, K. and Huang, H., 2018. Conditional generative adversarial network for gene expression inference. Bioinformatics, 34(17), pp.i603-i611.
39. Abu-Srhan, A., Abushariah, M.A. and Al-Kadi, O.S., 2022. The effect of loss function on conditional generative adversarial networks. Journal of King Saud University-Computer and Information Sciences, 34(9), pp.6977-6988.
40. Yan, B. and Han, G., 2018. Effective feature extraction via stacked sparse autoencoder to improve intrusion detection system. IEEE Access, 6, pp.41238-41248.