International Journal of Advanced Technology and Engineering Exploration, Vol 11(121)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2024.111100949

Research Article

Enhancing Arabic fake news detection with a hybrid MLP-SVM approach and Doc2Vec embeddings

Noralhuda Alabid[1]* and Hawraa Ali Taher[2]
Department of Computer Science, Faculty of Education, University of Kufa, Najaf, Iraq[1]
Department of Computer Science, Faculty of Education for Girls, University of Kufa, Najaf, Iraq[2]

Received: 02-June-2024; Revised: 08-December-2024; Accepted: 11-December-2024

Abstract
The rapid spread of the COVID-19 pandemic across the globe has resulted in the widespread dissemination of misinformation and fake news related to the virus. This misinformation has caused significant confusion among citizens, as well as heightened fear and anxiety. Consequently, it is crucial to develop automatic methods capable of effectively detecting such misinformation. Various machine learning (ML) and deep learning (DL) approaches have been implemented, each with its strengths and weaknesses. Combining multiple approaches has shown promising results. In this study, a hybrid model combining a multi-layer perceptron (MLP) and a support vector machine (SVM) was proposed to detect Arabic fake news using the ArCOVID19-Rumors dataset. The MLP was used to extract relevant features, while the SVM was employed to make the final classification decision. The proposed method also leverages Doc2Vec technology, a widely used document embedding technique, to extract features and convert them into numerical vectors while preserving the semantic and syntactic information of the documents. The experiment demonstrated that the hybrid MLP-SVM model outperformed related models when implemented independently. The results showed that the hybrid model achieved an accuracy of 87%, surpassing the performance of standalone MLP (84%) and SVM (83%) models. These results were validated using multiple metrics, including precision, recall, F1-score, and accuracy. The study indicates the importance of combining multiple algorithms for detecting fake news, as integrating the strengths of different techniques can lead to significantly improved classification performance.

Keywords
Fake news detection, Hybrid model, Multi-layer perceptron (MLP), Support vector machine (SVM), Doc2Vec embeddings, Arabic text classification.

1.Introduction
Social media plays a significant role in disseminating information. However, it also facilitates the spread of misinformation and fake news, which can have harmful impacts on health, finances, and politics, as evidenced by events like the U.S. Presidential election [1] and Brexit [2]. Numerous explanations and definitions attempt to capture the meaning of fake news. Farhoudinia et al. [3] defined fake news as "news articles that were deliberately launched to mislead the reader." The authors further classify fake news into two categories: disinformation (false news intentionally distributed to mislead) and misinformation (false news shared without the intent to mislead) [3].

During the outbreak of COVID-19, social media such as Twitter, Facebook, and Instagram played a significant role in circulating pandemic news globally. Some nefarious entities have exploited these platforms to spread fake news for various purposes.

As a result, numerous studies have focused on detecting fake news, particularly during the COVID-19 pandemic [4–6]. Additionally, specialized fact-checking websites, such as the Anti-Rumours Authority [7] and the Washington Post Fact Checker [8], were established to verify the validity of news. While these platforms are generally reliable, their effectiveness has diminished due to their limited capacity to manage the vast volume of news generated daily. Recently, the adoption of machine learning (ML) and deep learning (DL) techniques has gained prominence as a more scalable solution for addressing fake news [9, 10].

*Author for correspondence

Although this topic has been extensively researched, the majority of studies have focused on the English language. Detecting fake news in Arabic remains a relatively underexplored area. Processing Arabic content on social media presents unique challenges, including the prevalence of informal writing styles, frequent spelling errors, diverse dialects, an extensive vocabulary, complex morphological patterns, and the limited availability of Arabic datasets. These challenges have contributed to the lack of research specifically targeting the detection of Arabic fake news. Most existing approaches rely on traditional ML or DL techniques [11–15].

Most studies on Arabic language processing have been trained to target specific topics in fake news posts, relying primarily on the semantic meaning of individual words while overlooking the relationships between words that contribute to broader contextual understanding. This limitation creates a gap in effectively comprehending the overall context of the text. Additionally, the performance of current detection methods still requires significant improvement.

To address these issues, a novel model was proposed that employs Doc2Vec, which captures the meaning of individual words within their context, producing richer text representations. These representations are then integrated into a hybrid model combining a multi-layer perceptron (MLP) and a support vector machine (SVM). This hybrid model leverages the feature extraction capabilities of MLP and the classification strengths of SVM, offering an innovative approach to detecting fake news in Arabic.

To evaluate the effectiveness of the MLP-SVM hybrid model, performance metrics such as precision, accuracy, recall, and F1-score were calculated for the hybrid model, as well as for MLP and SVM individually. The results were compared to assess the improvements achieved by the hybrid approach. Furthermore, these results were also benchmarked against other baseline studies using the same ArCOVID19-Rumors dataset and another dataset called iSarcasm dataset to demonstrate the model's robustness and generalizability.

The main contributions of this study are as follows:

- To develop a unique hybrid approach that integrates Doc2Vec, MLP, and SVM to enhance fake news detection in Arabic.
- To enhance document representation by utilizing Doc2Vec to capture both the semantic meaning of words and their contextual relationships within the broader text, ensuring a more comprehensive understanding of the content.
- To integrate MLP and SVM by combining the feature extraction capabilities of MLP with the classification strength of SVM, resulting in a more robust and efficient model.
- To validate performance by evaluating the effectiveness of the proposed model and comparing its results with existing methods using the same dataset, thereby highlighting its superiority.

The rest of the paper is organized as follows: Section 2 reviews the related literature; Section 3 describes the methods of the proposed system; Section 4 provides details of the experiments and presents the results; Section 5 discusses the findings; and Section 6 concludes the study.

2.Literature review

The field of verifying false information on the web is a growing area of research that requires a combination of different approaches to achieve the best results. Wu et al. [16] noted that the process of web misinformation checking is a growing area of research. Thus, it requires a combination of different approaches to achieve optimal results. Recently, significant efforts have been made to handle the problem of spreading fake news, rumors, misinformation on social media. These efforts included the usage of many techniques from ML algorithms such as supervised methods, unsupervised methods and DL methods [17, 18]. Garg et al. [19] introduced a hybrid model combining convolutional neural network (CNN) and SVMs to detect fake news. The CNN was used for features extraction, while the SVM was utilized for further classification. This integration leveraged the strengths of CNN and SVM to enhance the accuracy results, achieving an accuracy of up to 88%. However, this model struggles with the increased complexity of computation and the longer training times. Dev and Bhatnagar [20] proposed hybrid SVMs and random forest (RF) algorithms for improving evaluation metrics of detecting fake news. The researchers showed that their model outperformed other traditional ML classifiers, achieving an accuracy of 97.56% and an F1-score of 93.50%. This model is characterized by simplicity and fast training times. However, it has difficulty capturing semantic relationships in text, particularly when handling

complex language structures. Detecting Arabic fake news faces challenges due to the complexity of the Arabic language, which varies significantly across regions and dialects [21]. This complexity makes it difficult to define and develop a comprehensive set of features suitable for use in ML models [22]. To overcome these challenges, researchers have explored various techniques for detecting Arabic fake news, including ML methods, DL models, and hybrid approaches. Abd et al. [23] optimized a combination model involving three tasks: multi-task learning, a transformer-based model, fire hawk optimizer algorithm (FHO), to address the problem of identifying fake news in Arabic. The use of fire hawk optimizer enhances the accuracy of the model. FHO is an effective algorithm that employs global and local techniques to find the best options. However, it is computationally complex, which indicates the need for improvements to reduce training time and increase its applicability in real-world applications. Himdi et al. [24] introduced a ML model to evaluate the credibility of Arabic news articles. They collected realistic news articles related to a specific domain. Then, the acquired dataset was used to construct fake news articles based on crowdsourcing. They extracted linguistic features, including sentiment, grammar, polarity, and parts of speech. The obtained features were trained by using classifiers like naive Bayes (NB), RF, and SVM classifiers to detect Arabic fake news. Their results showed that the extracted linguistic features were effective in identifying fake news. The best results were achieved using RF, which obtained an accuracy of 0.79 on the dataset. Bahurmuz et al. [25] proposed two methods for Arabic rumor detection: a BERT-based model pre-trained on Arabic text (AraBERT) and a BERT-based model specifically trained for modern Arabic (MARBERT). Their obtained results showed that relying on the transformer outperformed other DL models. However, this model showed low performance when trained on general datasets. This result highlights the issue of domain dependency. The best performance of this model relied on resampling techniques, such as oversampling, to address class imbalance. However, this approach may increase the risk of bias or overfitting, potentially affecting the model's generalizability. Hawashin et al. [26] developed an optimized feature selection method to identify Arabic fake news. Chi-square feature selection was applied to enhance the performance of a fake news detector utilizing various ML classifiers including: Logistic Regression (LR), SVM, RF, k-nearest neighbors, (KNN), NB, and AraBERT. Results demonstrated that applying Chi-square for feature section improved the accuracy of traditional ML methods. However, the efficiency of Chi-square feature selection based on detecting the optimal number of features that could vary across various dataset. Najadat et al. [27] proposed using various DL techniques to distinguish whether a news headline is actually linked or not to a parallel news article. They used both of long short-term memory (LSTM) and CNN with dataset related to the conflict in Syria and Middle East arguments. This dataset consists of 422 claims and 3,042 articles, which is relatively small and insufficiently to address the diversity and complexity of Arabic fake news. Harrag and Djahli [28] extended the [27] work by developing automatic fact-checking models for Arabic fake news-based CNN. This model was extensively evaluated to checking its ability to classify fake Arabic news. As reported, the model yielded and accuracy 91 % which outperformed state-of-the-art models used the same Arabic dataset. However, despite the improvement in results, the study also faced limitations due to the small dataset size, which restricted the generalizability of the results. Ameur and Aliane [29] believed that the efficiency of the fake news model depends on the size of dataset. Thus, they created a large manually classified dataset related to Arabic fake news about COVID-19. They used their dataset to train three pre-train transformer-based models to assist the efficiency of their dataset. Results showed that their dataset contributed to enhanced results compared to obtained results of the baseline models. Amoudi et al. [30] suggest a comparative research to investigate the performance several ML and DL models for fake new detection. They trained their model by using the ArCOV19-Rumors dataset. They conducted two investigations. In the first phase, they examined the effect of each of n-gram, TF-IDF, and word2vec on two ensemble learning methods which was voting and stacking as well as some of ML classifiers. The results confirmed outperform of stacking classifiers (accuracy of 0.81) than all other classifiers. In the second phase, they explored the influence of utilizing various optimizers with some of DL models. They found that the LSTM and Bidirectional LSTM (BiLSTM with the root mean square propagation (RSMprop) optimizer obtain the best accuracy results of 0.80 among all the other neural networks. Alyoubi et.al [31] utilized both of news content and social context to construct a DL model (CNN and BiLSTM) to detect fake news in Arabic Tweets. Their model was implemented with various word embedding models which are: Word2Vec, FastText, ARBERT, and MARBERT. The results showed that MARBERT

with CNN achieved the best value in accuracy and F1-score of 0.95. However, the imbalance of fake news tweets compared to real ones was the main challenge of this model. Another recent study [32] utilized a combination of (Text-CNN) and (LSTM) architectures to create an effective approach for classifying truth tweets in Arabic news. Fouad et al. [33] design a hybrid model consisted from first component, which is either AraBERT, GigaBIT, or MARBERT, and the second components, which was either a one-dimension CNN (1D-CNN) or a two-dimensional CNN (2D-CNN). The experiments conducted based on three datasets namely: Arabic news stance (ANS), Arabic news dataset (AraNews), and Covid-19-Fakes. The best results achieved were F1-scores of 0.6188 on the ANS dataset. However, ANS dataset suffer from data imbalance where the number of fake new is much smaller than real news. This imbalance can negatively impact on model generalization. Aljamel et al. [34] introduced a comparative study between two types of neural network models: recurrent neural network (RNN)-based models and BERT-based models. The RNN-based models included simple RNN, LSTM, gated recurrent unit (GRU), Bidirectional LSTM, and Bidirectional GRU, while the BERT-based models included AraBERT, ARBERTv2, CamelBERT, and QaribBERT. The results highlighted the efficiency of CamelBERT (achieved an accuracy of 0.86). This performance indicated the capability of CamelBERT to capture complex linguistic patterns and semantic relationships. The main obstacle of this model is its high computational demands, which require significant time and resources to explore additional hyperparameters to improve its performance. A recent research study published in [35] presented a transformer-based approach for detecting Arabic fake news. In their methodology, the authors explored various ensemble approaches, including a weighted-average ensemble, hard voting, and soft voting. The obtained results demonstrated the effectiveness of ensemble models in improving performance. However, despite the significant boost brought by ensemble-based approaches, combining multiple datasets into one large dataset can bias the model toward specific datasets. This study [36] investigated Arabic fake news detection by integrating FastText word embedding with ML and DL methods. it also utilized various transformer-based models such as BERT, XLNet, and RoBERTa. Among all the tested models, the hybrid Bi-GRU-Bi-LSTM model demonstrated the best performance, achieved an accuracy of 0.98 and an F1 score of 0.98. However, the effectiveness of this model may decline when handling informal language. In addition, the complexity of DL models, such as Bi-LSTM and Bi-GRU, requires intensive computational operations, limiting their accessibility for all users.

Reviewed studies indicate that Arabic fake news detection models have progressed through the application of diverse approaches, including traditional ML, DL models, transformer-based methods, and hybrid architectures. Techniques such as CNN-SVM combinations and transformer-based models have shown notable improvements in accuracy and performance. However, significant challenges persist, such as domain dependency, computational complexity, imbalanced datasets, and the limited availability of Arabic datasets.

This analysis highlights the need for more comprehensive models capable of handling larger and more balanced datasets to effectively address the complexities of Arabic fake news detection. Furthermore, hybrid models that combine multiple techniques have not been sufficiently explored, particularly in the context of the Arabic language.

In this paper, a novel hybrid architecture was proposed, leveraging the feature extraction capabilities of MLP and the robust classification strengths of SVM. To the best of our knowledge, this was the first study to implement such a hybrid model for fake news detection in Arabic, addressing a critical gap in existing research.

3.Materials and methods

3.1Embedding technique in modelling

Document embedding is a natural language processing (NLP) technique that converts a document, represented as a sequence of words, into a numerical vector representation. This vector captures essential semantic and syntactic information about the document, making it suitable for various NLP tasks such as document classification, clustering, and similarity measurement.

Recent advancements in embedding techniques, including Word2Vec, Doc2Vec, global vectors for word representation (GloVe), and BERT, have significantly enhanced NLP performance by transforming words and documents into meaningful vector representations. These vectors serve as inputs for training models, such as traditional machine learning algorithms and deep neural networks, resulting in improved outcomes for a wide range of NLP tasks.

The proposed model utilizes the Doc2Vec method for document vector representation. Doc2Vec is an extension of the well-known Word2Vec algorithm. While Word2Vec and GloVe map individual words to vectors and ignore word order or document-level context, Doc2Vec extends this functionality by embedding entire documents. In the Doc2Vec algorithm, all words in the training dataset are embedded as vectors using the Word2Vec model, represented as columns in a matrix M. Additionally, each document is transformed into a vector representation stored in a separate matrix D. The document and word vectors are then concatenated or averaged to predict the next word in the context of the given document. BERT, on the other hand, is characterized by its ability to capture deep contextual relationships between words within a sentence by employing bidirectional training. While BERT is highly powerful and effective for capturing nuanced relationships, it can be computationally expensive and potentially excessive for tasks with smaller datasets or less complex requirements.

The motivation for using Doc2Vec in this study lies in its simplicity and its ability to preserve both structural and semantic relationships within a document. By embedding documents with similar content close together in the vector space, Doc2Vec is particularly effective for tasks such as document retrieval and summarization.

Moreover, Doc2Vec captures the semantic composition of documents by representing individual words along with their contextual relationships. This enables a richer and more nuanced understanding of the text, allowing the model to grasp themes, sentiments, and topics effectively. Such capabilities significantly enhance the model's effectiveness in detecting fake news by enabling it to discern the nuanced usage of the Arabic language [37–39].

3.2The multilayer perceptron (MLP)

The MLP is an artificial neural network commonly used for implementing supervised learning algorithms, such as classification. A typical MLP structure consists of multiple layers: an input layer, one or more hidden layers, and an output layer. Each layer is connected to both the preceding and subsequent layers, forming a fully connected network. The input layer receives the features of the data to be processed. The hidden layers, situated between the input and output layers, perform complex computations on the data, learning patterns and relationships between features. The output layer applies the necessary activation function to produce predictions or classifications based on the learned patterns. Data flows through the MLP in a feedforward manner, moving from the input layer to the output layer without any looping back. All layers (except the input layer) use a nonlinear activation function, enabling the network to model complex, non-linear relationships. The MLP is trained using supervised learning, optimizing the network's parameters based on a labeled training dataset. This training process enables the MLP to generalize and predict outcomes for new, unseen data. The MLP model is highly flexible, allowing customization of the architecture, such as the number of hidden layers, the number of nodes in each layer, and the choice of activation functions. It also supports various optimization techniques for training, including gradient descent and stochastic gradient descent. MLP can be implemented using several programming languages and libraries, with Python being a popular choice. Libraries such as scikit-learn provide tools like the MLPClassifier class, which simplifies the implementation of MLP-based classifiers. This class allows for easy customization of the architecture and training parameters, making it a versatile tool for building MLP models [40].

3.3Support vector machines (SVMs)

SVMs are supervised machine learning algorithms designed for both classification and regression tasks. Initially created for binary classification problems, SVMs have been adapted to address multi-class classification challenges through various enhancements. The fundamental concept of SVMs is to identify a hyperplane in a high-dimensional space that separates data into distinct classes while maximizing the margin—the distance between the hyperplane and the nearest data points, known as support vectors, which significantly influence the hyperplane's position.

SVMs possess several advantages, including the ability to handle non-linearly separable data using kernel functions and scalability for large datasets. They also exhibit strong generalization performance and are less prone to overfitting compared to algorithms like decision trees. These features make SVMs a robust choice for a variety of machine learning applications.

3.4Stages of hybrid model

This subsection outlines the stages of the proposed classification model designed to address the task of Arabic fake news detection by leveraging the

strengths of MLP and SVM. The subsequent section details the architecture of the proposed model, focusing on the integration of MLP and SVM approaches. The basic architecture of the proposed model is illustrated in *Figure 1*.

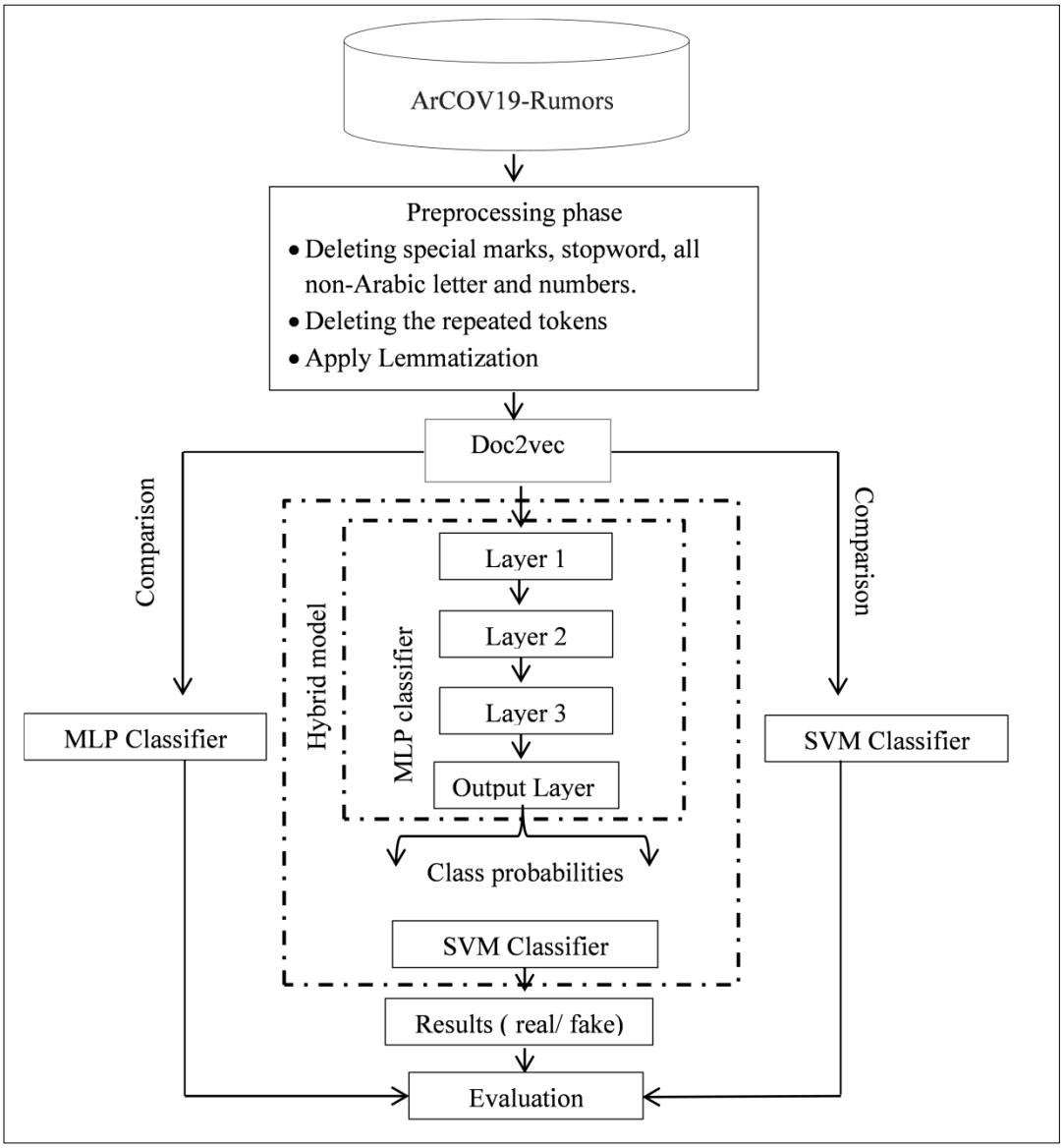


Figure 1 The flowchart of the proposed MLP-SVM model

3.4.1Dataset

The dataset used in this study is the ArCOV19-Rumors dataset [41], a specialized dataset collected to detect misinformation related to the COVID-19 pandemic. This dataset is a comprehensive Arabic Twitter dataset, spanning from January 27th, 2020, to January 31st, 2021, and containing approximately 2.7 million tweets. The dataset encompasses not only health-related content but also diverse claims influenced by COVID-19, covering topics such as social media issues, policy, sports, amusement, and religion. *Figure 2* illustrates some of the most frequently used words in the dataset. The ArCOV19-Rumors dataset is one of the first human-annotated Arabic Twitter datasets dedicated to the COVID-19 pandemic. It comprises two subsets:

- Claims subset: Contains tweets labeled as 'true,' 'false,' or 'other.'
- Tweet verification subset: Includes tweets confirming or refuting the original claims.

In the experiments, a subset of the claims data was used, consisting of 1,753 tweets labeled as false and 1,831 labeled as true. This subset was collected from 20 Arabic-speaking countries. The claims subset includes three columns: ID, text, and a label column, where "false" denotes fake information and "true" denotes factual information. No additional features were used, so NLP techniques were applied exclusively to the textual content for analysis and classification.

Figure 2 Words clouds of ArCOV19-Rumors

3.4.2Pre-processing

Several preparation steps were performed to validate the texts for the classification models. The dataset was pre-processed to convert the texts into a suitable format for classification models. The Natural Language Toolkit (NLTK) package was used to implement the pre-processing steps. Several techniques, outlined below, were applied to clean the dataset, which comprised a mixture of various Arabic dialect texts.

- Identifying hashtags and replacing them with related tokens to reduce noise in the dataset and improve the model's ability to understand the discussions topic.
- Deleting special marks, stop words, numbers, and all non-Arabic tokens and links which are not contributed to identify the meaning of the content and can confuse the model. This helps the model to focus on the more relevant words in texts and reducing distractions during the learning process.
- Deleting repeated tokens characters help address common issues in informal texts. This step allows the model to process these words as a single entity and accrue better generalization. This enables model to handle informal writing more effectively.
- Lemmatization was applied to the tokens to reduce words to their base forms; for example, "running" was converted to "run." This process helps reduce the size of the dataset and improves the model's ability to recognize and classify relevant features accurately.

Python was used to implement the programming tasks of pre-processing with utilizing Fastai library to perform Lemmatization and segmentation, and the NLTK library for other pre-processing steps.

3.4.3Dataset split and validation

To assess the proposed system's performance, the experiment was conducted using the ArCOV19-Rumors dataset. This dataset is composed of 1,831 articles classified as fake and 1,753 articles identified as real information. After the pre-processing step, the dataset was divided into training dataset 80% and testing dataset 20% by using a specific technique called k-fold cross-validation. This method splits the dataset into k subset of partitions. One of whole partitions is located for testing while remains are used for training. The process is repeated k time to ensure that every data point is allocated for training and testing process. Based on experiments, a 10-fold cross-validation approach was employed. This means that the dataset was divided into 10 partitions. With each iteration, a different partition was detected for testing while the remaining nine partitions were used for training. The final results were calculated based on the average across all partition. Additionally, A stratification technique was applied in this process to ensure the proper distribution of real and fake news in each partition. This step is crucial to avoid the model from being skewed toward a specific class.

Although there is a noticeable difference between the number of data points for fake news and real news samples in the dataset. There are no specific strategies such as class weights or resampling to treatment this issue in this study. The proposed hybrid MLP-SVM model was trained on the row dataset to evaluate its robustness in handling imbalanced data without external interventions.

3.4.4Mapping words to vector by using Doc2Vec

In this approach, the Doc2Vec model was trained on a corpus of Arabic text to generate numerical vector representations, which were used as features for the MLP classifier. The classifier was trained on labeled data to map features to target labels. The effectiveness of this method depended on the quality of the Doc2Vec model and the role of pre-processing, both of which were critical for improving classifier performance. After pre-processing the training data and creating tagged documents using the TaggedDocument class from the gensim library, the

Doc2Vec model was trained with the Doc2Vec class from the same library.

To prevent overfitting and optimize performance, key parameters of the model were carefully tuned. The vector size was set to 100, striking a balance between capturing sufficient semantic details and maintaining computational efficiency. The number of iterations was set to 50 to ensure the model converged effectively without overfitting, while the number of epochs was adjusted to 40, which was empirically found to produce stable embeddings without overfitting the training data. A minimum word count of 2 was specified to include only words that occurred more than once in the dataset. These parameter values were determined through empirical testing by experimenting with various combinations to achieve the best model accuracy. This procedure aligns with the recommendations of [42], which emphasize the importance of fine-tuning parameters to effectively capture semantic nuances in text data. *Figure 3* illustrates the steps for obtaining the document vectors.

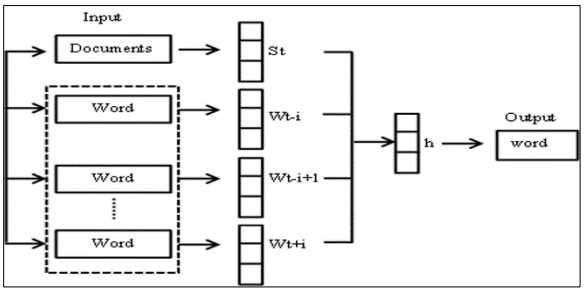


Figure 3 Document embedding framework

3.4.5Environment setup and parameter optimization

In this subsection, the overall framework of the proposed hybrid method for handling an Arabic corpus in the classification task is presented. *Figure 4* illustrates the main architecture of the proposed model, which explores the use of a hybrid MLP-SVM approach for classification. The integration of MLP and SVM demonstrated significantly improved performance compared to using MLP or SVM individually [43]. This improvement is attributed to the efficient handling of non-linear spaces, which enhanced classification accuracy by optimizing the adjustment of non-linear relationships within the features.

This hybrid approach allowed the model to better classify complex samples, making it more robust. Although the mixed method is slightly more complex, it resulted in notable improvements across various performance metrics, including accuracy, precision, recall, and F1-score. These enhancements are clearly evident when comparing the results of the hybrid model with those of the individual classifiers.
In the experiment, the Sklearn ML tool was used to apply MLP and SVM. Gensim library was used to call Doc2Vec representation in process of building the embedding vector. To perform these methods, an open-source Python libraries was used. Training was conducted using machine equipment with core i7, NVIDIA GPUs with 16 GB and 32 GB RAM.

The hyperparameters of the MLP and SVM models were set by trial and error strategic across several parameter configurations. For the MLP, experiments were conducted with different numbers of hidden layers which is ranging from 2 to 4, and neurons per layer (between 100 and 400). The final configuration was three hidden layers containing 400, 250, and 200 neurons respectively. These values were selected because it provided the best trade-off between performance and computation efficiency during cross-validation. To reduce overfitting, L2 regularization was applied with a parameter (alpha =0.02) after initial tests showed it improved generalization. Also, the ReLU activation function was utilized, coupled with the Adam optimizer for efficient training. Additionally, the maximum number of iterations was adjusted to 500 to ensure convergence within a reasonable time frame. The output of last layer of the MLP was then fed into the SVM to predict the genuine of news. For the SVM, after testing various kernels (polynomial, linear, and RBF), the polynomial kernel (kernel='poly') was identified as the most effective when combined with the MLP output. The parameter gamma was set to 'auto', and the maximum number of iterations was limited to 500 to ensure efficient model convergence. While the mixed method introduces a slightly higher level of complexity, it significantly improved key performance metrics, including accuracy, precision, recall, and F1-score. These improvements are evident when comparing the results of the SVM and MLP classifiers individually. Algorithm 1 provides a detailed explanation of the processes involved in the proposed model.

Algorithm 1: Hybrid MLP-SVM

Input: Dataset D Consisting of a set of articles A={a1,a2,...,an}

Output: Trained Hybrid MLP-SVM Model M

1. Data Pre-Processing

- *For each article ai ∈ A:*

1. Remove noise (e.g., white spaces, hashtags).
2. Remove articles a word count less than 3
3. Perform stop-word removal and Arabic stemming.
4. Tokenize each article and apply padding or truncation:
if len(tokens)<Θ, pad the article with zeros.
if len(tokens)> Θ, truncate it
5. Convert tokens into numeric values using encoding techniques.
• 18 Convert the text data into Doc2Vec embeddings
2. Doc2Vec Embedding
• Prepare Tagged Documents T={t1,t2,...,tn}
• Train a Doc2Vec model E on T:
Embedding vector vi=E(ti), ∀ti ∈ T
• Infer document embeddings for feature extraction.
3. Training the Hybrid MLP-SVM Model
• 3.1: Train a MLP classifier MMLP on vi with optimized hyperparameters:
MMLP ← argminθ L(y, MMLP(vi;θ)),
where L is the loss function and y are the labels
• 3.2: Train SVM classifier MSVM:
MSVM ← argmaxθ margin
• Combine the outputs of MMLP and MSVM for hybrid classification.
• Trained hybrid MLP-SVM model.
4. Prediction and Evaluation
• Predict labels ȳ using the hybrid model M: ȳ =M(vi)
• Evaluate performance using metrics: accuracy, precision, recall, and F1-score.
• End

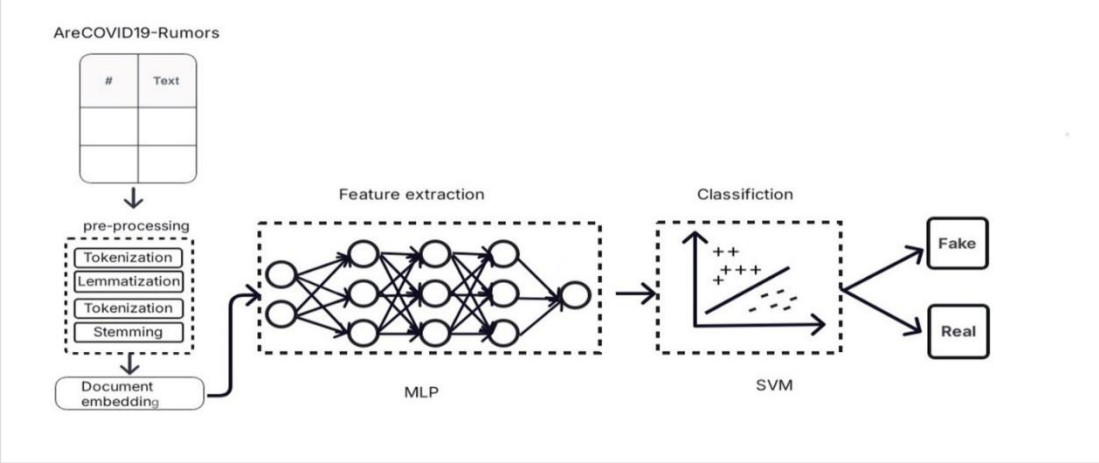


Figure 4 The overall architecture of the hybrid model

4.Results

4.1Evaluation metrics

Several measures were applied for evaluating results which were; accuracy, recall, and F1-score, and precision to evaluate systems as explained in Equations 1, 2, 3, and 4. The accuracy of the results is decided by the ratio of correctly selected samples to the total number of samples evaluated and classified. The recall is the proportion of the total positive samples that were correctly found. The F1-score is calculated by taking the harmonic mean of precision and recall, effectively balancing the two metrics. Precision is calculated by taking the total number of true positive results and dividing by the total number of positive results predicted by the model, including both true positives and false positives, as shown in Equations 1 to 4.

$$Accuracy = \frac{TruePositive + True\ negative}{TruePositive + True\ negative + False\ positive + False\ negative} \quad (1)$$

$$\text{Recall} = \frac{\text{True value}}{\text{TruePositive+FalseNegative}} \quad (2)$$

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision} * \text{recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

$$\text{Precision} = \frac{\text{True value}}{\text{True falue+False value}} \quad (4)$$

4.2Results based on MLP-SVM classifier

The results demonstrated that using the hybrid MLP-SVM classifier with Doc2Vec embeddings, combined with pre-processing of the Arabic texts,

significantly improved classification performance. The hybrid classifier achieved an accuracy of 87%, outperforming the standalone MLP (84%) and SVM (83%) classifiers. These results, summarized in *Table 1*, highlight the superior performance of the hybrid approach on Arabic text.

Table 1 Model performance with different classifiers

	Accuracy	F1-score	Recall	Precision
MLP-SVM	87	86	86	85
MLP	84	83	82	83
SVM	83	81	82	81

Additionally, there was a notable increase in precision, recall, and F1-score with the hybrid model. The hybrid MLP-SVM classifier achieved a precision of 85%, compared to 83% for MLP and 81% for SVM. Precision is crucial for evaluating the model's ability to accurately identify fake news without mistakenly labeling real news as fake.

The recall value also highlighted the hybrid model's strength in detecting fake news, achieving 86%, while MLP and SVM yielded 82% each. A high recall value ensures the model's capability to identify the largest possible number of fake news instances, thereby reducing the spread of false negatives.

Finally, the F1-score was calculated to assess the model's ability to balance the distinction between fake and real news. The hybrid model achieved an F1-score of 86%, compared to 83% for MLP and 81% for SVM. A higher F1-score indicates the model's effectiveness not only in detecting fake news but also in maintaining a balance between correctly identifying fake news (recall) and avoiding misclassification of real news as fake (precision).

Among all the models, the hybrid MLP-SVM model outperformed the others in its prediction process, as illustrated in the confusion matrix in *Figure 5*. The hybrid model accurately predicted 1,492 fake news items as fake. However, 209 news items were incorrectly classified as fake while they were actually real. Conversely, 1,346 real news items were correctly classified as real, while 232 real news items were incorrectly predicted as fake.

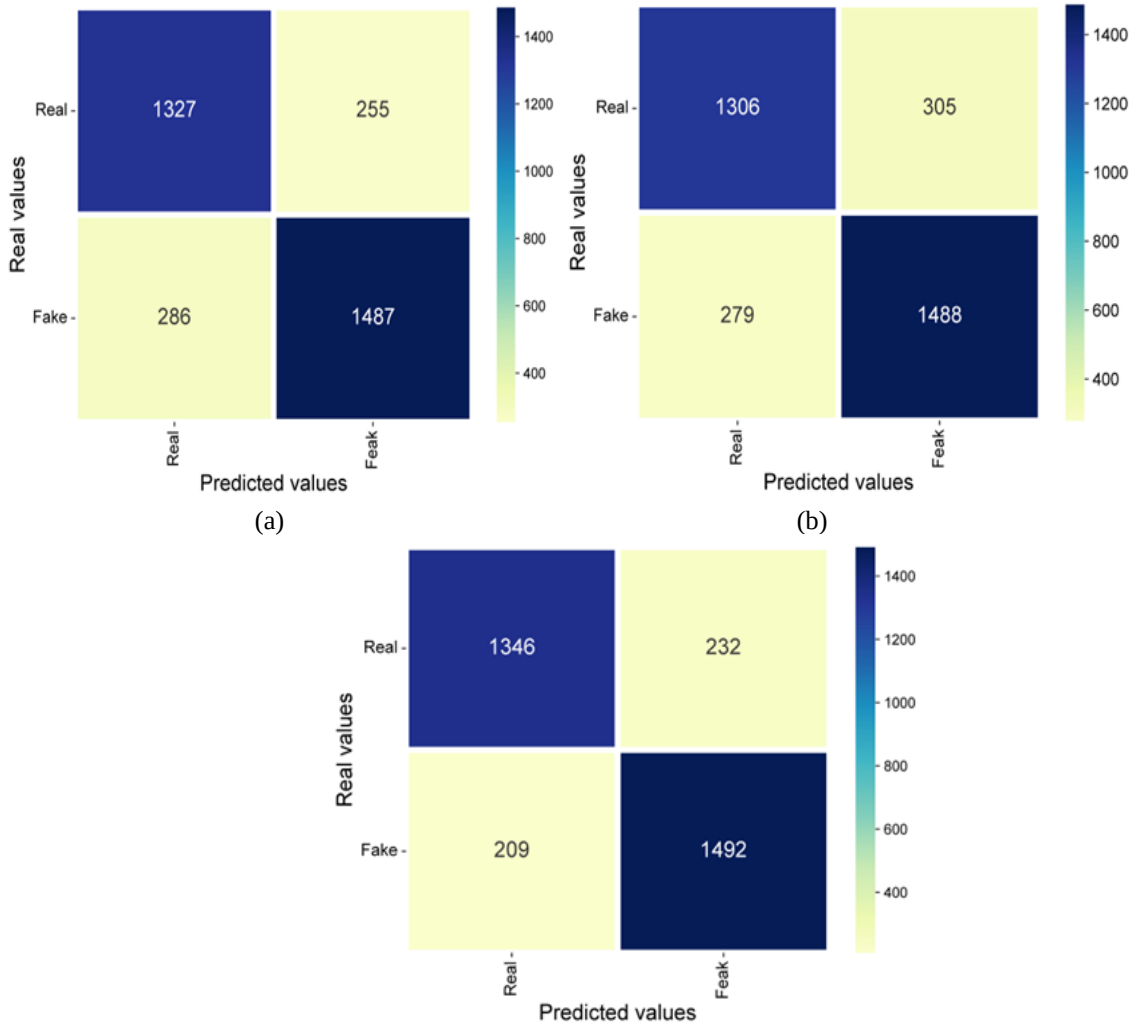
For further validation, the proposed hybrid model was compared with three previous models that utilized the ArCOV19-Rumors dataset [30, 31, 44]. Performance was evaluated based on accuracy and F1-score. The results clearly demonstrate that the proposed model outperformed the baseline models in terms of accuracy.

Additionally, a comparison was made with the outcomes from a ML model that used similar datasets [45]. *Table 2* presents the comparison results, showing that the proposed model achieved the best performance among all the evaluated models across two different datasets. This highlights the effectiveness and robustness of the hybrid MLP-SVM approach in detecting Arabic fake news.

Table 2 Provides a comprehensive comparison between the proposed model and other existing works

Models	Name of the dataset	Accuracy
MLP (Proposed model)	ArCOV19-Rumors	84
SVM (Proposed model)	ArCOV19-Rumors	83
MLP-SVM (Proposed model)	ArCOV19-Rumors	87
BiLSTM [30]	ArCOV19-Rumors	88
MARBERT-CNN [31]	ArCOV19-Rumors	86
GRU(WL-W2V) [44]	ArCOV19-Rumors	83.1
MLP [45]	iSarcasm dataset	83.6
SVM [45]	iSarcasm dataset	81

The ROC curve was applied to evaluate the results of the hybrid MLP-SVM model. It represents the relationship between the false positive rate (FPR) and the true positive rate (TPR) for the diagnostic test data. The ROC curve enables model assessment based on the value of the AUC (Area Under the Curve) variable, which quantifies the area under the plotted curve between these two dimensions. *Figure 6* shows the ROC curve for the proposed MLP-SVM model using Doc2Vec embeddings. To further explain the performance of the proposed model, especially in the presence of class imbalance, *Figure 7* presents the precision-recall curve for the MLP-SVM model. This curve highlights the balance between precision and recall, providing deeper insights into the model's ability to handle imbalanced data effectively.



(a) (b)

(c)

Figure 5 Confusion matrix: a) for MLP, b) for SVM, and c) for the hybrid model

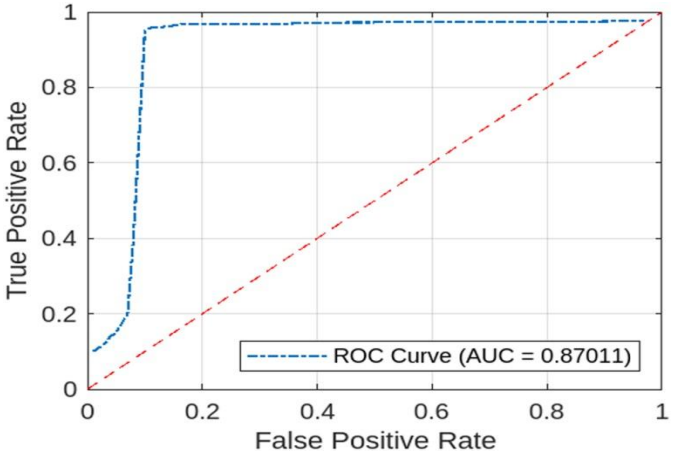


Figure 6 ROC curve of MLP-SVM

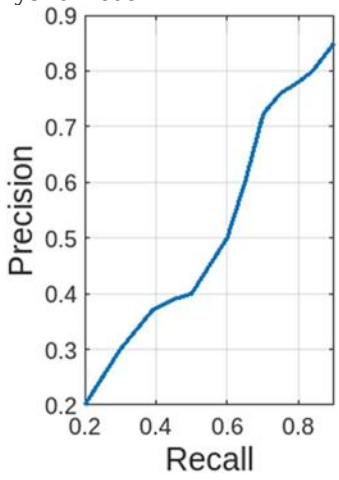


Figure 7 Precision-Recall curves of MLP-SVM

5.Discussion

The effectiveness of the hybrid proposed system was evaluated by comparing its performance with standalone MLP and SVM models. The hybrid system (MLP-SVM) demonstrated its robustness by outperforming MLP and SVM individually, achieving significantly higher accuracy, precision, recall, and F1-score. This improvement was attributed to the combination of feature extraction using Doc2Vec and the strengths of MLP and SVM.

Using Doc2Vec for feature extraction enabled the representation of Arabic documents as fixed-length vectors containing meaningful semantic content. These vectors provided informative features for classification when used as inputs to the hybrid model. The inclusion of semantic and contextual information allowed the proposed model to better differentiate between fake and real news in Arabic contexts. MLP effectively learned and extracted significant features through its multiple layers of neurons, capturing complex patterns and relationships in Arabic texts, including informal writing styles and diverse dialects. SVM, on the other hand, managed high-dimensional data by maximizing the margin between classes and minimizing overfitting. The combination of MLP's feature detection and SVM's classification robustness made the hybrid model particularly effective in detecting Arabic fake news. By integrating MLP and SVM, the hybrid model improved classification capabilities, with MLP detecting complex text features and SVM isolating fake news from real news. This collaboration allowed for a more comprehensive analysis of textual data, resulting in enhanced performance metrics.

When compared to previous baseline methods that used the same dataset (ArCOV19-Rumors), the proposed hybrid model demonstrated a clear advantage. It achieved an accuracy of 87.37%, surpassing the GRU (WL-W2V) model with 83.1% [44], the BiLSTM model with 80% [30], and the MARBERT-CNN model with 86% [31]. Additionally, the hybrid model was compared with a baseline study [45] that used a different dataset but employed MLP and SVM individually, achieving 83.6% accuracy for MLP and 81% for SVM. In contrast, the proposed model showed improved performance even with standalone MLP (84%) and SVM (83%), thanks to the use of Doc2Vec for feature selection. Moreover, the hybrid model consistently achieved higher precision, recall, and F1-score values, highlighting its superior ability to detect fake news in Arabic texts. The study's use of a hybrid MLP-SVM model represents a novel contribution, as none of the previous baseline studies combined these techniques.

Despite its promising results, the proposed model has some limitations. Combining MLP and SVM significantly increases training and computation time, particularly when handling large datasets, making scalability challenging or requiring resource optimization. Additionally, due to the sensitivity of MLP and SVM to hyperparameters, identifying the optimal combination of hyperparameters for both models can be time-consuming. The slightly imbalanced dataset used for training may also bias the model toward the majority class. Furthermore, the model is designed specifically for detecting fake textual news, limiting its applicability to other formats, such as images, videos, or multimedia content, reducing its effectiveness in such scenarios.

A complete list of abbreviations is listed in *Appendix I.*

6.Conclusion

The problem of fake news is not new. With the rise of the internet, false information spreads rapidly through social media platforms, posing significant risks to societies. Therefore, the ability to distinguish between fake and real news is crucial. This study addresses the issue of false news detection in Arabic tweets during the COVID-19 pandemic by leveraging the benefits of integrating MLP and SVM algorithms. These algorithms were combined to create a hybrid classifier aimed at enhancing classification performance. To ensure effective implementation, comprehensive preprocessing techniques were applied to the ArCOV-19 dataset. Additionally, k-fold cross-validation was employed to split the dataset into training and testing sets. The Doc2Vec embedding method was then used to extract features, enabling the capture of the semantic essence of text data and transforming tokens into meaningful vector representations. These vectors were passed through the layers of the MLP classifier to select the most relevant features and perform an initial classification. Subsequently, SVM classified the news articles as fake or genuine based on the features identified by the MLP. The integration of MLP and SVM into a hybrid classifier improved classification performance significantly. The results demonstrated that the hybrid classifier achieved higher accuracy rates compared to the individual classifiers, offering a promising approach to mitigating the spread of fake

news during the pandemic. For future work, several enhancements can be explored, such as using alternative embedding methods like GloVe, Word2Vec, FastText, and BERT. Additionally, integrating MLP with other machine learning models may provide further insights into performance improvements and yield more robust solutions.

Acknowledgment
None.

Conflicts of interest
The authors have no conflicts of interest to declare.

Data availability
The ArCOV19-Rumors dataset used in this study is publicly accessible and can be obtained from https://paperswithcode.com/paper/arcov19-rumors-arabic-covid-19-twitter.

Author's contribution statement
Noralhuda Alabid: Conceptualization, investigation, writing – original draft, writing – review and editing. **Hawraa Ali Taher:** Conceptualization, writing – original draft, analysis and interpretation of results.

References
[1] Raza S, Khan T, Chatrath V, Paulen-patterson D, Rahman M, Bamgbose O. FakeWatch: a framework for detecting fake news to ensure credible elections. Social Network Analysis and Mining. 2024; 14 (1):142.
[2] Maci S, Demata M, Mcglashan M, Seargeant P. The Routledge handbook of discourse and disinformation. Routledge; 2024.
[3] Farhoudinia B, Ozturkcan S, Kasap N. Fake news in business and management literature: a systematic review of definitions, theories, methods and implications. Aslib Journal of Information Management. 2023:1-24.
[4] Chen MY, Lai YW, Lian JW. Using deep learning models to detect fake news about COVID-19. ACM Transactions on Internet Technology. 2023; 23(2):1-23.
[5] Alsmadi I, Rice NM, O'brien MJ. Fake or not? automated detection of COVID-19 misinformation and disinformation in social networks and digital media. Computational and Mathematical Organization Theory. 2024; 30(3):187-205.
[6] Xia H, Wang Y, Zhang JZ, Zheng LJ, Kamal MM, Arya V. COVID-19 fake news detection: a hybrid CNN-BiLSTM-AM model. Technological Forecasting and Social Change. 2023; 195:122746.
[7] Liu Y, Wu YF. Fned: a deep network for fake news early detection on social media. ACM Transactions on Information Systems. 2020; 38(3):1-33.
[8] Al-sarem M, Alsaeedi A, Saeed F, Boulila W, Ameerbakhsh O. A novel hybrid deep learning model for detecting COVID-19-related rumors on social media based on LSTM and concatenated parallel CNNs. Applied Sciences. 2021; 11(17):1-17.
[9] Capuano N, Fenza G, Loia V, Nota FD. Content-based fake news detection with machine and deep learning: a systematic review. Neurocomputing. 2023; 530:91-103.
[10] Villela HF, Corrêa F, Ribeiro JS, Rabelo A, Carvalho DB. Fake news detection: a systematic literature review of machine learning algorithms and datasets. Journal on Interactive Systems. 2023; 14(1):47-58.
[11] Touahri I, Mazroui A. Survey of machine learning techniques for Arabic fake news detection. Artificial Intelligence Review. 2024; 57(6):157.
[12] Azzeh M, Qusef A, Alabboushi O. Arabic fake news detection in social media context using word embeddings and pre-trained transformers. Arabian Journal for Science and Engineering. 2024:1-4.
[13] Abd EDS, Abdelaziz A, Essam G, Mohamed SE. AraFake: a deep learning approach for Arabic fake news detection. In international mobile, intelligent, and ubiquitous computing conference 2023 (pp. 1-8). IEEE.
[14] Alotaibi T, Al-dossari H. A review of fake news detection techniques for Arabic language. International Journal of Advanced Computer Science & Applications. 2024; 15(1):392-407.
[15] Dahou A, Ewees AA, Hashim FA, Al-qaness MA, Orabi DA, Soliman EM, et al. Optimizing fake news detection for Arabic context: a multitask learning approach with transformers and an enhanced nutcracker optimization algorithm. Knowledge-Based Systems. 2023; 280:111023.
[16] Wu J, Guo J, Hooi B. Fake news in sheep's clothing: robust fake news detection against LLM-empowered style attacks. In proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining 2024 (pp. 3367-78).
[17] Alghamdi J, Luo S, Lin Y. A comprehensive survey on machine learning approaches for fake news detection. Multimedia Tools and Applications. 2024; 83(17):51009-67.
[18] Hu L, Wei S, Zhao Z, Wu B. Deep learning for fake news detection: a comprehensive survey. AI Open. 2022; 3:133-55.
[19] Garg N, Gupta R, Kaur M, Ahmed S, Shankar H. Efficient detection and classification of orange diseases using hybrid CNN-SVM model. In international conference on disruptive technologies 2023 (pp. 721-6). IEEE.
[20] Dev DG, Bhatnagar V. Hybrid RFSVM: hybridization of SVM and random forest models for detection of fake news. Algorithms. 2024; 17(10):1-16.
[21] Sabri T, El BO, Kissi M. Comparative study of Arabic text classification using feature vectorization methods. Procedia Computer Science. 2022; 198:269-75.
[22] Nassif AB, Elnagar A, Elgendy O, Afadar Y. Arabic fake news detection based on deep contextualized embedding models. Neural Computing and Applications. 2022; 34(18):16019-32.

[23] Abd EM, Dahou A, Orabi DA, Alshathri S, Soliman EM, Ewees AA. A hybrid multitask learning framework with a fire hawk optimizer for Arabic fake news detection. Mathematics. 2023; 11(2):1-15.
[24] Himdi H, Weir G, Assiri F, Al-barhamtoshy H. Arabic fake news detection based on textual analysis. Arabian Journal for Science and Engineering. 2022; 47(8):10453-69.
[25] Bahurmuz NO, Amoudi GA, Baothman FA, Jamal AT, Alghamdi HS, Alhothali AM. Arabic rumor detection using contextual deep bidirectional language modeling. IEEE Access. 2022; 10:114907-18.
[26] Hawashin B, Althunibat A, Kanan T, Alzu'bi S, Sharrab Y. Improving Arabic fake news detection using optimized feature selection. In international conference on information technology 2023 (pp. 690-4). IEEE.
[27] Najadat H, Tawalbeh M, Awawdeh R. Fake news detection for Arabic headlines-articles news data using deep learning. International Journal of Electrical & Computer Engineering. 2022; 12(4): 3951-9.
[28] Harrag F, Djahli MK. Arabic fake news detection: a fact checking based deep learning approach. Transactions on Asian and Low-Resource Language Information Processing. 2022; 21(4):1-34.
[29] Ameur MS, Aliane H. AraCOVID19-MFH: Arabic COVID-19 multi-label fake news & hate speech detection dataset. Procedia Computer Science. 2021; 189:232-41.
[30] Amoudi G, Albalawi R, Baothman F, Jamal A, Alghamdi H, Alhothali A. Arabic rumor detection: a comparative study. Alexandria Engineering Journal. 2022; 61(12):12511-23.
[31] Alyoubi S, Kalkatawi M, Abukhodair F. The detection of fake news in Arabic tweets using deep learning. Applied Sciences. 2023; 13(14):1-21.
[32] Wotaifi TA, Dhannoon BN. An effective hybrid deep neural network for Arabic fake news detection. Baghdad Science Journal. 2023; 20(4):1392-401.
[33] Fouad KM, Sabbeh SF, Medhat W. Arabic fake news detection using deep learning. Computers, Materials & Continua. 2022; 71(2):1-19.
[34] Aljamel A, Khalil H, Aburawi Y. Comparative study of fine tuned BERT-based models and RNN-based models. case study: Arabic fake news detection. The International Journal of Engineering & Information Technology. 2024; 12(1):56-64.
[35] Al-zahrani L, Al-yahya M. Pre-trained language model ensemble for Arabic fake news detection. Mathematics. 2024; 12(18):1-7.
[36] Ealmandouh M, Alrahmawy MF, Eisa M, Elhoseny M, Tolba AS. Ensemble based high performance deep learning models for fake news detection. Scientific Reports. 2024; 14(1):1-24.
[37] Elsaeed E, Ouda O, Elmogy MM, Atwan A, El-daydamony E. Detecting fake news in social media using voting classifier. IEEE Access. 2021; 9:161909-25.
[38] Mitroi M, Truică CO, Apostol ES, Florea AM. Sentiment analysis using topic-document embeddings. In 16th international conference on intelligent computer communication and processing 2020 (pp. 75-82). IEEE.
[39] Rhanoui M, Mikram M, Yousfi S, Barzali S. A CNN-BiLSTM model for document-level sentiment analysis. Machine Learning and Knowledge Extraction. 2019; 1(3):832-47.
[40] Geeitha S, Aakash R, Akash G, Arvind AM, Thameem AS, Mahudapathi P, et al. Enhanced artificial neural network for spoof news detection with MLP approach. In international conference on advanced communications and machine intelligence 2022 (pp. 441-51). Singapore: Springer Nature Singapore.
[41] Haouari F, Hasanain M, Suwaileh R, Elsayed T. ArCOV19-Rumors: Arabic COVID-19 twitter dataset for misinformation detection. In proceedings of the sixth Arabic natural language processing workshop 2012 (pp.72-81).
[42] Dogru HB, Tilki S, Jamil A, Hameed AA. Deep learning-based classification of news texts using doc2vec model. In 1st international conference on artificial intelligence and data analytics 2021 (pp. 91-6). IEEE.
[43] Siamidoudaran M, İşçioğlu E. Injury severity prediction of traffic collision by applying a series of neural networks: the city of London case study. Promet-Traffic & Transportation. 2019; 31(6):643-54.
[44] Al-yahya M, Al-khalifa H, Al-baity H, Alsaeed D, Essam A. Arabic fake news detection: comparative study of neural networks and transformer-based approaches. Complexity. 2021; 2021(1):5516945.
[45] Ashraf N, Elkazzaz F, Taha M, Nayel H, Elshishtawy T. BFCAI at SemEval-2022 task 6: multi-layer perceptron for sarcasm detection in Arabic texts. In proceedings of the 16th international workshop on semantic evaluation (SemEval-2022) 2022 (pp. 881-4).

Noralhuda Alabid is a lecturer and researcher in the Department of Computer Science at the Faculty of Education, University of Kufa (UOK), located in Najaf, Iraq. She has been a faculty member at UOK since 2012 and holds an MS in Advanced Computer Science. She earned her undergraduate degree in Computer Science from the Faculty of Science at the University of Sheffield in the UK. Her academic background is in computer science, with expertise in image processing, bioinformatics, natural language processing, text processing, data mining, pattern recognition, and medical statistics.
Email: noralhudan.hadi@uokufa.edu.iq

Hawraa Ali Taher graduated from the Department of Computer Science and Mathematics at Kufa University, Iraq, with a B.Sc. degree in Computer Science in 2009 and an M.Sc. degree in Computer Science in 2020. She is currently employed as an Assistant Teacher at Kufa University. Her research interests include Artificial Intelligence and Natural Language Processing.
Email: hawraa.alshimirty@uokufa.edu.iq

Appendix I

S. No.	Abbreviation	Description
1	1D-CNN	One-Dimensional Convolutional Neural Network
2	2D-CNN	Two-Dimensional Convolutional Neural Network
3	ANS	Arabic News Dataset
4	AraBERT	A BERT-based model pre-trained on Arabic text
5	ARBERT	A pre-trained BERT model fine-tuned for Arabic
6	BERT	Bidirectional Encoder Representations from Transformers
7	BiLSTM	Bidirectional Long Short-Term Memory
8	CamelBERT	A transformer-based model fine-tuned for Arabic
9	CNN	Convolutional Neural Network
10	Covid19Fakes	A dataset related to fake news on COVID-19 in Arabic
11	Doc2Vec	Documents to vectors
12	F1-score	The harmonic mean of precision and recall
13	FastText	A word embedding model
14	FHO	Fire Hawk Optimizer
15	GloVe	Global Vectors for Word Representation
16	GRU	Gated Recurrent Unit
17	KNN	k-nearest neighbors
18	LR	Logistic regression
19	LSTM	Long Short-Term Memory
20	MARBERT	A BERT-based model specifically trained for modern Arabic
21	ML	Machine Learning
22	n-gram	a sequence of n adjacent symbols in particular order
23	NLP	Natural language processing
24	NLTK	Natural Language Toolkit
25	NVIDIA GPUs	Graphics Processing Units made by NVIDIA
26	QaribBERT	Qarib Bidirectional Encoder Representations from Transformers.
27	ReLU	Rectified Linear Unit
28	RF	Random Forest
29	RNN	Recurrent Neural Network
30	RoBERTa	Robustly Optimized BERT Pretraining Approach
31	ROC	Receiver Operating Characteristic
32	RSMprop	Root Mean Square Propagation
33	SVM	Support Vector Machine
34	Text-CNN	A text classification model using Convolutional Neural Networks
35	TF-IDF	Term Frequency-Inverse Document Frequency.
36	Word2Vec	Word to Vector
37	XLNet	eXtreme Language Model Network