



An Effective Hybrid Model for Fake News Detection in Social Media Using Deep Learning Approach

R. Raghavendra¹ · M. Niranjnamurthy²

Received: 15 June 2023 / Accepted: 12 February 2024
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2024

Abstract

Social media is a strong Internet platform that allows people to voice their thoughts about numerous events happening in real time at multiple locations. People comment and express their thoughts on any social media post. Meanwhile, fake news or misleading information is disseminated regarding exciting occurrences that occur in real time. A large number of Internet users read and distribute such false and fake material without knowing the nature or legitimacy of the news. This has a detrimental influence on people's perceptions of the particular event. Classical techniques are utilized to determine the type of news distributed on Twitter, Instagram, YouTube, and Facebook. However, these techniques fail to take into account characteristics such as news-generating location, consistency, timing, and novelty. It eventually leads to a scenario in which individuals form incorrect ideas and have misleading perceptions regarding any startling news. The spread of misleading thoughts and remarks has a significant impact on real-world action results. This research article addresses these difficulties by developing a system for detecting fake news using deep learning techniques and models. To determine the originality of the news, place of generation, and longevity, two deep learning models are developed artificial neural network and a mixed classifier model of convolution neural networks and long short-term memory. This aids in detecting bogus news and removing it from the server where it is stored. The tests utilizing these coupled models increase the detection of false information in social media. Furthermore, geo-map is used in this research to aids in the regulation of fake news flowing on social media regarding unique occasions occurring all over the world.

Keywords Artificial neural network (ANN) · Convolutional neural network (CNN) · Long short-term memory (LSTM) · False/fake news detection (FND) · Natural language processing (NLP) · Stemming

Introduction

The amount of social media data used on an internet platform and accessed by end users is increasing on a daily basis in the digital world. People in the real world use social media to communicate their ideas and opinions about current events. Users are rarely concerned about the originality and veracity of data generated on an internet site. As a result, bogus news and incorrect remarks spread quickly on social media. Previously, such phony news was discovered using a variety of methods. Using traditional methodologies, fake news propagation is identified based on data, features, models, and applications [1]. In terms of temporal, dataset, and psychological methodologies, the 'data concerned' approach is utilized to identify fake news. Data mining methods are used to classify the findings. The 'feature concerned' technique works on news material and social environments on the internet platform. The findings are shown and compared

This article is part of the topical collection "Diverse Applications in Computing, Analytics and Networks" guest edited by Archana Mantri and Sagar Juneja.

✉ R. Raghavendra
raghuramesh88@gmail.com

M. Niranjnamurthy
niruhdsd@gmail.com

¹ Department of MCA, BMS Institute of Technology and Management (Affiliated to Visvesvaraya Technological University), Jnana Sangama, Avalahalli, Belgavi, Bangalore, Karnataka 560064, India

² Department of AI and ML, BMS Institute of Technology and Management (Affiliated to Visvesvaraya Technological University), Jnana Sangama, Avalahalli, Belgavi, Bangalore, Karnataka 560064, India

to see if there is room for improvement in recognizing fake news. The third method, model concerned, employs machine learning techniques such as supervised, semi-supervised, and unsupervised learning to detect fake news distributed on social media. Diffusion and intervention approaches identify bogus news in every event reported on social media using the ‘application-concerned’ method [2–5]. All approaches identify news from social media based on evaluations such as favorable or negative, good or terrible, and the keywords used by users. The approaches for detecting bogus news on social media are depicted in Fig. 1.

With many parameters and criteria, classic techniques such as data mining, the Naive Bayes theorem, support vector machines (SVM), random forests, and logistic regression are used to forecast the spread of fake news on social platforms [6–10]. However, because of the flood of communications published on the internet and the rising number of individuals accessing social media platforms, the accuracy of fake news identification remains a challenge. This results in unfavorable or negative ideas and thoughts about a certain incident, which are then shared on social media. The neural network is utilized at several layers to operate on such information on social media. Classifiers are used in this neural network technology to identify unwanted and nonsensical data. With sophisticated neural network ideas such as CNN, these classifiers construct a model that is used to detect fake news [11–15]. It is also being developed to detect bogus news by adding layers to it. The following tiers of CNN use specialized circuits called long short-term memory (LSTM) to record the changes in news that are available on social media [16–21]. On a real-time system, our approach detects fake news using classifiers from social media data. Nonetheless, the data status changed largely by social media users at each second, potentially flopping to hold all the statuses consistently. To solve problems in social media, hybrid models of CNN and LSTM classifiers are utilized, which give more accuracy than previous models.

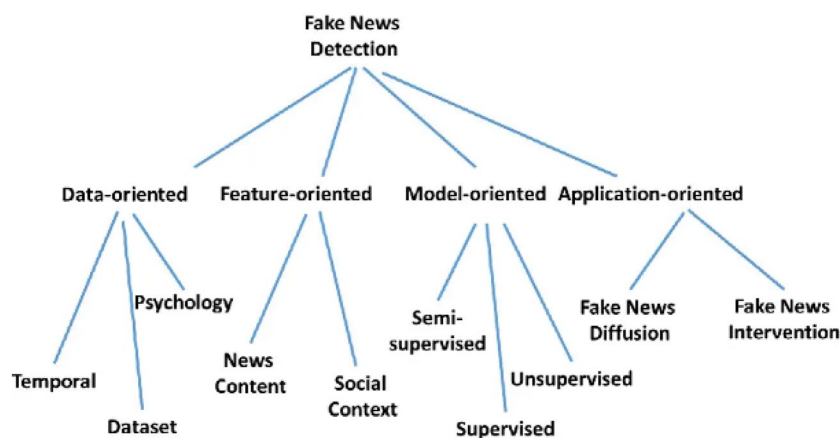
Literature Review

Hanshal et al. proposes a hybrid deep learning model for automatic fake news detection, addressing the escalating threat of misinformation on social media. The model combines convolutional neural network (CNN) and recurrent neural network (RNN), integrating an automatic data augmentation method, auxiliary classifier generative adversarial networks. Focusing on news headlines and content, the model achieves superior results with 93.87% accuracy, 10.39% recall, and 93.12% precision on various datasets. The approach involves preprocessing steps, stop word removal, and blockchain use, ensuring robust classification performance while mitigating noise and overfitting. The study emphasizes the significance of data enhancement in maintaining classifier efficiency.

Jing et al. introduces a novel multimodal progressive fusion network (MPFN) for fake news detection, addressing the limitations of existing semantic-based methods. MPFN employs a transformer structure for visual feature extraction, progressively sampling features at different levels and fusing information from text and image frequency domains. The model achieves superior accuracy, outperforming other state-of-the-art methods by at least 4.3%, particularly on Twitter data. By integrating deep and shallow information from diverse modalities, MPFN demonstrates advanced fake news identification capabilities, mitigating infodemic challenges. The proposed approach effectively establishes fine-grained correlations between modalities, showcasing its efficacy in multimodal disinformation detection.

Rawat et al. addresses the pervasive issue of fake news through a machine learning-based approach. Utilizing the TF-IDF feature extraction method and support vector machine (SVM) classifier, the proposed system achieves a 100% recognition rate on the designed datasets. The study identifies key features for effective fake news detection, prioritizing text, author, source, date, and sentiment. N-gram

Fig. 1 Types of fake news detection methods



encoding surpasses the bag of words method, and SVM proves to be the optimal algorithm. The paper suggests future work should involve a larger dataset and explore online learning for continuous updates in detecting evolving fake news.

Lai et al. used classical approaches like binomial logistic regression, Naive Bayes (NB), support vector machines (SVM), random forests (RF), neural networks, convolutional neural networks (CNN), long short-term memory networks (LSTM), CNN with global maxpools, CNN with deep networks, and LSTM for false news detection with the help of the frequency-inverse document frequency (TF-IDF) method. The detection of fake news is done with the Word2Vec method, which exposes the vectorization of the words present in the news from internet sources, especially social media. The exact word can be taken from each sentence, and their vector values are identified using that model in order to remove the extra information added in the original news. Frequency-inverse document frequency (TF-IDF) is identified as the repeated or selected word for categorizing to predict the matches.

Mridha et al. proposed new models, namely BERT (Bidirectional Encoder Representations), BDANN (BERT-based domain adaptation neural network), GCN (graph convolutional network), and GNN (graph neural network), with the following techniques: attention, generative adversarial networks, and bidirectional encoder representations for transformers to detect fake news. Each and every model is used to detect the fake news that is expressed in social media results immediately with the help of their nature types. All models are based on deep learning techniques, which use many layers to predict the news or comments effectively.

Nasir et al. developed a new hybrid model with RNN (recurrent neural networks) and CNN techniques to detect fake news automatically with the benchmark data sets ISO and FA-KES effectively. The layers of these techniques detect false information in the social media from the pre-trained model as a combined technique of RNN and CNN.

Khan et al. [7] discussed a pre-trained model for fake news classification with the help of an advanced BERT model along with benchmark datasets. The length of the dataset and its diversity are processed for prediction to explore the range of the data. The advanced models of BERT are used for further developments.

Kaliyar et al. developed a system to detect fraudulent activities on social media using CNN and LSTM models with the PHEME COVID-19 dataset. This method reveals the false information entered in the PHEME dataset, which contains the details of patients all over the world. Based on the LSTM circuit model classifier, its originality and time of generation were identified, and actual data were only available on the social media platform for trusted statistics.

LSTM keeps track of the data status throughout the session and their liveliness time in the platform. It helps to prevent the spread of fake news on social media from different locations around the world.

Zeng et al. proposed a pre-trained VGG model that contains variation multimodal encoders for detecting fake news. Meanwhile, they proposed that FND use the semantic correlations between text content and images attached (FND-SCTI) model to differentiate the original and false news from the text and images. This method classifies both fake text and images from social media and pulls them out periodically. The datasets used to perform this method are from Twitter and Weibo. This dataset contains both text and images as a combination of positive and negative comments.

Aslam et al. discussed an ensemble-based deep learning model to avoid the spread of fake news on online platforms with BiLSTM (bidirectional LSTM) techniques. It stores the previous and next states of the data accessed on the platforms in order to predict the liveliness of the original data. LIAR is considered for fake news detection using BiLSTM. Ashik et al. narrated the model called BiLSTM to develop the fake news detection process effectively. It works with deep learning concepts and holds the status of the data in both directions, like previous and next turn.

Chauhan et al. proposed a model based on a deep learning LSTM neural network with NLP concepts. Initially, the text as a piece of news is considered for word embedding using the gloVe tool and performing vectorization to predict the occurrence of words. The tokenization technique applies those words, then their features are extracted for further development. Finally, the N-grams technique is used for predicting fake news from the collection of words on the social media platform. This model developed fake news detection without the latest algorithms but only with data structures and data mining principles.

Choudhary et al. discussed the fake news spreading avoidance concepts with news embedding block (NEB) and multi-scale feature block (MSFB) techniques. Here the familiar classifier BERT is used, and their advanced model BerConvoNet is involved in finding the false information in online data. The layers of BerConvoNet extract information from social media and do analysis with their models using deep learning techniques. A framework classifies all the news available on social media and analyzes it immediately to predict fake news.

Sadr et al. and Seyam et al. discussed the previous models such as KNN, SVM, Naive Bayes, CNN, and the advanced models of LSTM and BiLSTM techniques for detecting fake news classification. Both models are used to perform FND based on the time of their lives and the rate at which their states change rapidly. Chitra et al. [16], Lee et al. [22], Drif et al. [23], and Ramya et al. [24] developed models with CNN, LSTM, and deep learning techniques using LIAR

data sets. Also, models of SVM, Nave Bayes, Adaboost, CNN, and GRU (gated recurrent unit) with the benchmark datasets LIAR, Buzzfeed, Politifact, Kaggle, and McIntire are discussed. The entire system develops the design of the unit to predict fake news with the above techniques, which processed several datasets. While comparing the results from the different datasets, it seems like DL models produce more accurate results than the previous methods.

Islam et al. and Umer et al. summarized their models with the combinations of CNN, RNN, LSTM, BiLSTM, and GRU to remove the false information on social media. All techniques use deep learning techniques to predict in real time. In addition, BERT, XLNet, and RoBERT models are also used to improve the accuracy of the results. PCA (Principal Component Analysis) and chi-square techniques are used for detecting fake images from internet sources. Detecting fake images is a challenging concept in social media for a huge number of end-users. This model has been successfully developed to perform those operations in real time.

Kaliyar et al., Aniyath et al., and Zhang et al. developed models with deep learning techniques along with CNN and LSTM through the LIAR dataset. It was developed fully on the attention-based concepts that are used in DL techniques. The results are compared with the previous models, and their accuracy levels increased when they used DL models. The more layers used in this model are an advantage over the previous techniques. The features of the text and images were classified when the dataset contained both values to predict the originality of the contents in social media.

Justification for Research Gap

Research Gap 1: Lack of Comparative Analysis on Hybrid Models

While Hanshal et al. present a robust hybrid model for fake news detection, there is a notable research gap in the absence of a thorough comparative analysis with other existing hybrid models. The paper highlights the effectiveness of their convolutional neural network (CNN) and recurrent neural network (RNN) combination, along with auxiliary classifier generative adversarial networks (AC-GAN) for data augmentation. However, a comprehensive evaluation against alternative hybrid architectures, such as those incorporating transformer-based models or attention mechanisms, is missing. Addressing this gap would contribute to a deeper understanding of the strengths and weaknesses of various hybrid approaches, aiding researchers and practitioners in selecting the most suitable models for specific contexts.

Research Gap 2: Limited Exploration of Cross-Modal Fusion Techniques

Jing et al.'s multimodal progressive fusion network (MPFN) introduces an innovative approach to fake news detection by combining text and image information. However, the research gap lies in the limited exploration of cross-modal fusion techniques. While the paper demonstrates the effectiveness of the transformer structure for visual feature extraction and progressive sampling, it does not thoroughly investigate alternative fusion mechanisms, such as attention-based fusion or cross-modal attention. Exploring and justifying the choice of fusion methods would enhance the understanding of how different modalities contribute to fake news detection and could potentially uncover more efficient fusion strategies.

Research Gap 3: Lack of Scalability Analysis in SVM-based Approach

Rawat et al.'s approach using the TF-IDF feature extraction method and support vector machine (SVM) classifier achieves impressive results, with a 100% recognition rate. However, the research gap here is the absence of a scalability analysis. The study does not provide insights into how the proposed system's performance scales with larger datasets, and it does not address potential challenges when dealing with real time or continuously evolving fake news scenarios. Justifying the scalability and adaptability of the SVM-based approach to varying data sizes and continuous updates is essential for ensuring the practicality and long-term applicability of the proposed fake news detection system.

Proposed Work

Using convolutional long short-term memory (CLSTM) networks in fake news classification provides advantages in capturing spatial dependencies and extracting features from textual content. The combination of convolutional and recurrent operations enables efficient processing of local patterns, reducing computational complexity and leading to faster training times. CLSTMs' adaptability to varying input sizes is particularly beneficial for handling news articles of different lengths commonly found in real-world datasets. This model architecture demonstrates improved generalization capabilities, contributing to its effectiveness in classifying fake news by capturing underlying patterns. Overall, CLSTMs offer a robust framework that leverages both convolutional and recurrent layers, making them well-suited for the complexities of fake news classification tasks.

The proliferation of fake news in society underscores the critical need for robust detection mechanisms, and this

project proposes a hybrid deep learning model designed for the identification of fake news within the realm of social media.

The system's architecture shown in Fig. 2 incorporates two primary models: a hybrid convolutional neural network (CNN) and long short-term memory (LSTM) model, alongside an artificial neural network (ANN) model for benchmarking purposes. The workflow encompasses key stages, commencing with the training of the system on a meticulously labeled news article dataset distinguishing between real and counterfeit information. Natural language processing techniques are subsequently applied to preprocess the news articles, encompassing text cleaning, stop word removal and word stemming or lemmatization. The preprocessed text is then transformed into dense vector representations through an embedding layer, capturing the underlying semantic meanings. The dataset is divided into training, validation, and test sets to facilitate model training, validation, and evaluation. Two models are constructed: an ANN model well-suited for classification tasks and a CLSTM model tailored for sequential data classification, such as text. These models are individually trained using the training data and subsequently evaluated on the test set, with the results meticulously analyzed to ascertain the superior performer in the task of fake news detection. The hybrid approach amalgamates the strengths of CNN and LSTM models, offering an enhanced and comprehensive solution to combat the dissemination of misinformation in social media and its detrimental societal repercussions.

NLP Stemming for Preprocessing

During the data preprocessing step, data sets from online sources are collected using identified keys associated with the complete story. The stemming technique is used to remove the additional letters attached at the end of each

term and label them in the correct sequence. Similarly, all feasible letters are sorted, and the leftover characters are readily deleted using NLP. The final value of the news is saved as a data set. This is the input to the hybrid model that is used to detect fake news. If there are more words in the paragraph, the label system is utilized to determine the order of the news. Symbols, digits (0–9), and special characters with additional components are eliminated. Figure 3 depicts the stemming and dataset development processes in NLP.

Dataset

The dataset under consideration comprises a compilation of articles classified as either real or fake news. The real news articles were meticulously sourced from Reuters.com, a reputable and widely recognized news outlet, ensuring a foundation of credibility. In contrast, the fake news articles were curated from diverse origins identified by both Politifact and Wikipedia as disseminators of misinformation. This dual-sourcing strategy aims to create a comprehensive and balanced dataset, reflective of both trustworthy and questionable information. The dataset covers an array of subjects, yet a predominant emphasis is placed on political and global news. By incorporating articles from these varied sources and focusing on key thematic areas, the dataset provides a robust foundation for studying the nuances and patterns associated with real and fake news, facilitating research and analysis in the realm of information authenticity and media literacy.

This study utilizes a dataset sourced from a fake news competition on Kaggle. The objective of the competition is to develop a system capable of discerning unreliable news articles. The dataset comprises text snippets and metadata related to news articles and is organized into two files: train.csv and test.csv. More than 1,40,000 records were in this dataset.

Fig. 2 Proposed architecture

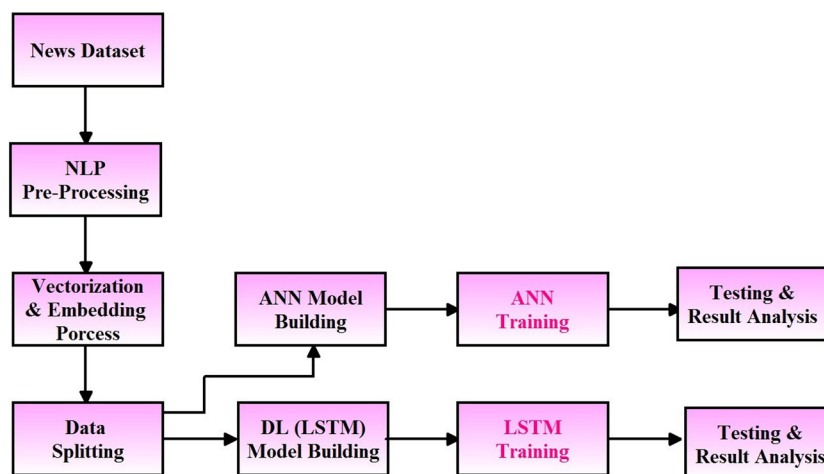


Fig. 3 NLP-stemming for data preprocessing

	original_word	stemmed_words
0	connect	connect
1	connected	connect
2	connection	connect
3	connections	connect
4	connects	connect

	original_word	stemmed_word
0	trouble	troubl
1	troubled	troubl
2	troubles	troubl
3	troublesome	troublesom

changing
changed
change

stemming

chang
chang
chang

studying
studies
study

stemming

studi
studi
studi

train.csv: This file serves as a comprehensive training dataset and includes the following attributes:

- id: a unique identifier for each news article
- title: the title of the news article
- author: the author of the news article
- text: the content of the article, which may be incomplete
- label: a binary indicator categorizing the article as potentially unreliable
 - 1: unreliable
 - 0: reliable

test.csv: This file functions as a testing dataset, mirroring the attributes found in train.csv, excluding the label field.

substantial data for effective training. Overfitting, where the model learns the training data too well and struggles with new data, is a potential challenge. To mitigate this, techniques like dropout and regularization are employed.

Deep learning, a subfield of machine learning, has gained prominence largely due to the success of deep neural networks, which are ANNs with many hidden layers. This allows them to automatically discover intricate features from data, leading to state-of-the-art performance in various domains including computer vision, speech recognition, and natural language processing.

Despite their complexity, ANNs are a cornerstone of modern machine learning and have led to significant advancements in fields such as healthcare, finance, and autonomous systems, revolutionizing how we process and understand complex data (Fig. 4).

Artificial Neural Network (ANN)

Artificial neural networks (ANNs) are computational models inspired by the structure and functioning of the human brain. They consist of interconnected nodes or “neurons” organized in layers. The input layer receives data, which is then processed through hidden layers, and finally, an output layer produces the result. Each connection between neurons has an associated weight, which adjusts during training to optimize predictions. ANNs excel at learning complex patterns in data, making them widely used in tasks like image recognition, natural language processing, and regression analysis. They possess the ability to generalize from training data to make predictions on unseen data.

One of the key strengths of ANNs lies in their ability to perform nonlinear transformations on data. This means they can learn and model complex relationships that may not be evident through traditional statistical methods. However, ANNs can be computationally intensive and require

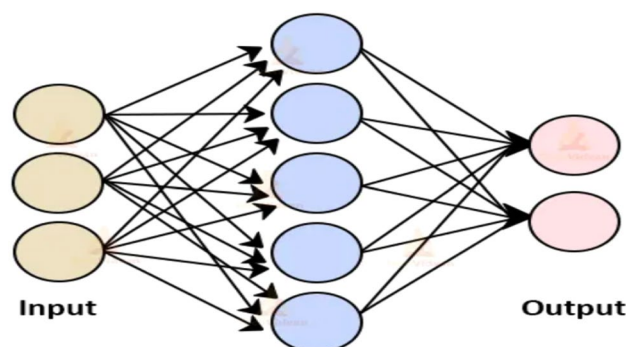


Fig. 4 Artificial neural network architecture

Convolutional Long Short-Term Memory (CLSTM)

Convolutional long short-term memory networks (CLSTMs) represent a fusion of two powerful neural network architectures: convolutional neural networks (CNNs) and long short-term memory networks (LSTMs). This combination enables CLSTMs to effectively process grid-like structured data, such as images or time series. CNNs are excellent at automatically extracting relevant features from input data. They employ filters or kernels to scan over the data, capturing spatial hierarchies of features. This makes them especially well-suited for tasks like image recognition. LSTMs, on the other hand, are proficient in modeling sequential or time-dependent patterns. They are equipped with a gating mechanism that allows them to selectively retain and forget information over long sequences, making them adept at tasks where context and order are critical.

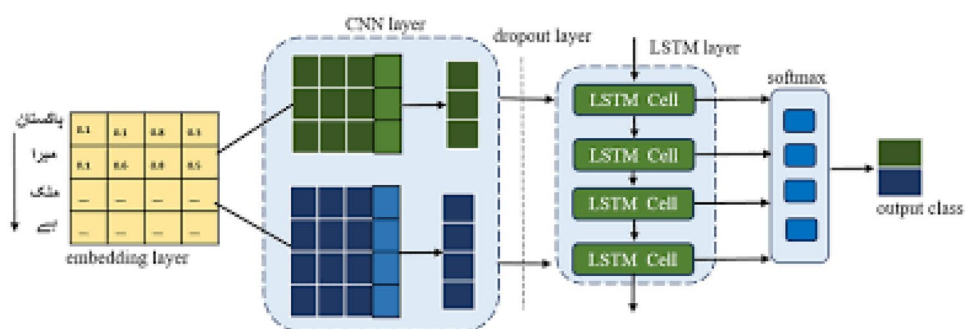
By incorporating both convolutional layers and LSTM units, CLSTMs bring the best of both worlds to the table. They can learn spatial features from data, which is crucial for tasks like object detection in images. Simultaneously, they can capture long-term dependencies and temporal patterns, which is vital for tasks where understanding sequences of data is essential, like in video analysis or natural language processing.

CLSTMs have found success in various domains including video action recognition, weather forecasting, and medical image analysis. They leverage the local information processing of CNNs and the memory retention of LSTMs to make accurate predictions on data with both spatial and sequential characteristics. This adaptability and versatility make CLSTMs a powerful tool in the field of deep learning, particularly for tasks involving complex data with intricate spatiotemporal relationships (Fig. 5).

Algorithm 1 CLSTM classification process

1. Prepare the data for CLSTM model: ``X_train, X_test, y_train, y_test = split_data(F)``
2. Define the CLSTM model architecture: ``CLSTM_model = define_CLSTM_model()``
3. Train the CLSTM model: ``CLSTM_model.train(X_train, y_train)``
4. Evaluate the model on the test data: ``accuracy = CLSTM_model.evaluate(X_test, y_test)``
5. Get the predicted class: ``predicted_class = CLSTM_model.predict(F)``
6. Output the predicted class: ``print("Predicted class:", predicted_class)``

Fig. 5 Convolutional long short-term memory architecture



The CLSTM classification process begins with data preparation, where the features are split into training and testing datasets, namely X_{train} , X_{test} , y_{train} , and y_{test} . Subsequently, a convolutional long short-term memory (CLSTM) model is defined, tailored for fake news classification. The model is then trained using the training dataset (X_{train} , y_{train}) to learn relevant patterns and relationships. Its performance is assessed by evaluating accuracy on the test data (X_{test} , y_{test}). Once trained, the CLSTM model can predict the class of news based on their features, facilitating efficient fake news classification system.

Implementation & Result Analysis

Implementation of artificial neural networks (ANNs) involves setting up layers, defining activation functions, and configuring training parameters. ANNs excel at tasks like image recognition and regression analysis. Result analysis entails evaluating metrics like accuracy, precision, and recall to gauge model performance. Convolutional long short-term memory networks (CLSTMs) combine CNNs and LSTMs for tasks involving grid-like data, such as images or time series. Implementation includes structuring layers for both spatial and sequential processing. CLSTMs excel in video analysis and sequential data tasks. Result analysis involves

Fig. 6 ANN model summary

Model: "sequential"		
Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 50, 300)	10317300
batch_normalization (Batch Normalization)	(None, 50, 300)	1200
dropout (Dropout)	(None, 50, 300)	0
dense (Dense)	(None, 50, 16)	4816
batch_normalization_1 (Batch Normalization)	(None, 50, 16)	64
dense_1 (Dense)	(None, 50, 32)	544
batch_normalization_2 (Batch Normalization)	(None, 50, 32)	128
dropout_1 (Dropout)	(None, 50, 32)	0
flatten (Flatten)	(None, 1600)	0
batch_normalization_3 (Batch Normalization)	(None, 1600)	6400
dense_2 (Dense)	(None, 64)	102464
batch_normalization_4 (Batch Normalization)	(None, 64)	256
dropout_2 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 2)	130
=====		
Total params: 10433302 (39.80 MB)		
Trainable params: 10429278 (39.78 MB)		
Non-trainable params: 4024 (15.72 KB)		
=====		

assessing model predictions, often using techniques specific to sequential data, like time series analysis or video frame-level evaluation. Comparing performance across these architectures helps determine the most suitable approach for a given task.

ANN Model Summary

The artificial neural network (ANN) model comprises layers of interconnected neurons. Each layer processes data, and weights between neurons adjust during training. It excels at learning complex patterns (Fig. 6).

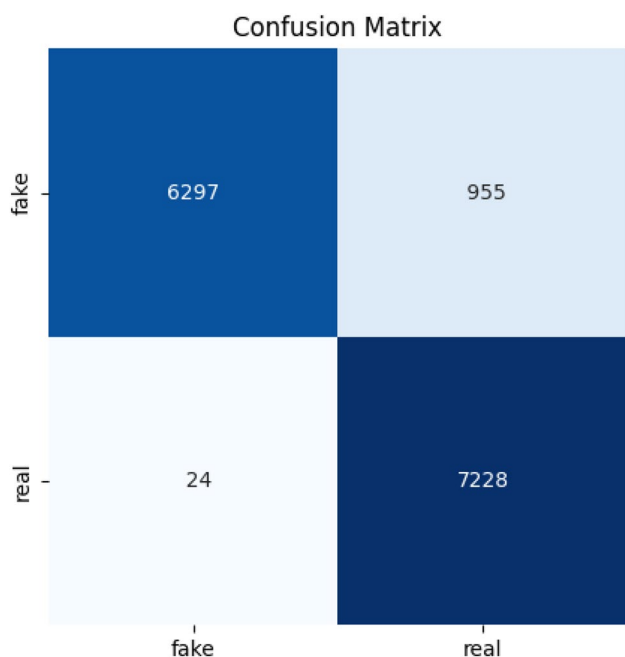
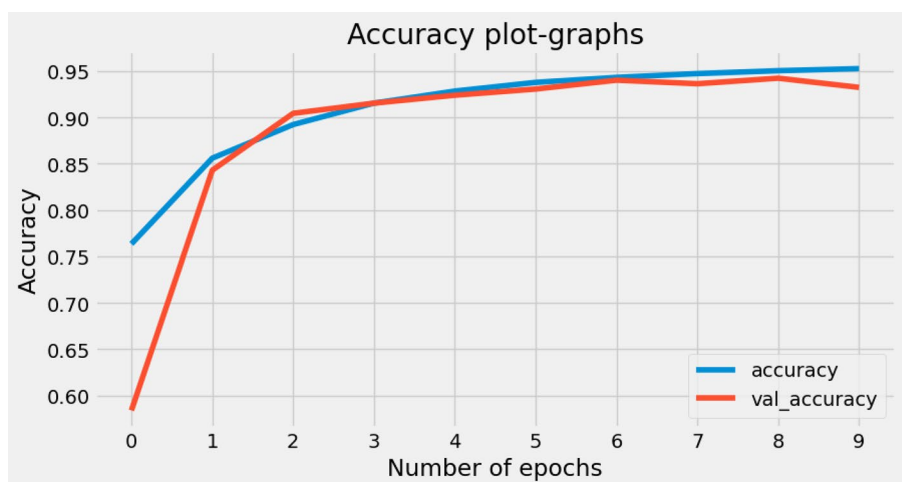


Fig. 7 Confusion matrix for ANN

Fig. 8 Accuracy plot graph



The provided Python code snippet utilizes the Keras library to define a neural network model, tailored for text classification tasks. The model architecture is sequentially constructed, with layers added in succession. Beginning with an embedding layer, it transforms integer-encoded tokens into 300-dimensional vectors using pre-trained word embeddings specified by the **'embedding_matrix'**. Batch normalization is then applied to stabilize activations, followed by dropout regularization to mitigate overfitting. The subsequent layers consist of densely-connected units with ReLU activation functions, promoting effective feature learning. Additional batch normalization and dropout layers are interspersed to further enhance model robustness. The **'Flatten'** layer is introduced to prepare the input for subsequent fully connected layers. The architecture culminates in an output layer with a sigmoid activation function, indicative of a binary classification problem. The model is compiled with the binary cross-entropy loss function, well-suited for binary classification tasks, along with the Adam optimizer for efficient gradient descent. Accuracy is chosen as the metric to monitor during training. This comprehensive structure demonstrates a sophisticated neural network design, carefully incorporating embedding, normalization, and dropout techniques to address the challenges inherent in text classification. The model's intricacy is a testament to its potential for capturing intricate relationships within textual data, essential in tasks requiring nuanced understanding and classification of written content.

Confusion Matrix-ANN

In Fig. 7, the confusion matrix for the ANN model is presented. A close analysis of the matrix reveals outstanding performance. Among the 14,504 total samples, the model accurately identifies 6297 as fake and misclassifies 955. Similarly, for real targets, 7228 are correctly predicted,

with only 24 misclassifications. This culminates in an impressive accuracy rate of 93%.

Accuracy & Loss Plot Graph for ANN

An accuracy plot graphically illustrates a model's performance across iterations or epochs during training. It displays accuracy scores (y-axis) against iteration/epoch numbers (x-axis). Initially, accuracy may vary, reflecting the model's struggle with complexity. With more iteration, accuracy generally improves, indicating better generalization. A plateau may occur, indicating diminishing returns. Monitoring this plot guides decisions on when to stop training to prevent overfitting and offers insights into the model's learning curve for fine-tuning and real-world optimization (Fig. 8).

Loss Plot Graphs

A loss plot graph illustrates a machine learning model's performance during training. It plots the loss function value on the y-axis against the number of epochs on the x-axis. Initially, the high loss indicates significant prediction errors. With training, the loss diminishes, indicating improved accuracy. The ideal graph displays a consistent decline, though minor fluctuations are normal. A sudden spike in loss may suggest overfitting, where the model fits too closely to the training data and performs poorly on new data. Conversely, a stagnant loss may indicate the model has plateaued in learning. A well-constructed loss plot is vital for comprehending training progress and making informed model adjustments (Fig. 9).

CLSTM Model Summary

The CLSTM model combines convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) for processing grid-like data with spatial and sequential characteristics, enabling tasks like video analysis (Fig. 10).

The provided Python code defines a sophisticated neural network model using Keras, designed for text classification tasks. The architecture begins with an embedding layer, converting integer-encoded tokens into dense vectors with 300 dimensions, leveraging pre-trained word embeddings indicated by the 'embedding_matrix'. Subsequently, two consecutive 1D convolutional layers with 64 filters and ReLU activation functions are added, enabling the model to capture local patterns. Batch normalization and max pooling layers follow to stabilize activations and enhance feature extraction. This is succeeded by another pair of convolutional layers with 128 filters each, further refining the model's ability to detect complex patterns. Batch normalization and max pooling are again applied to consolidate the learned features. The introduction of bidirectional LSTM layers with 100 units and return sequences facilitates the capture of long-range dependencies in the data. This bidirectional architecture allows the network to consider context from both past and future information. A flattening layer is then introduced to prepare the data for subsequent dense layers. Dropout regularization with a rate of 0.5 is applied to reduce overfitting. Two dense layers with 256 and 'len(class_labels)' units, activated by ReLU and sigmoid functions respectively, conclude the architecture. The model is compiled with a binary cross-entropy loss function, the Adam optimizer, and accuracy as the chosen metric for training evaluation. This intricately designed architecture adeptly combines embedding, convolutional, recurrent, and dense layers, showcasing a nuanced approach to extracting and comprehending

Fig. 9 Loss plot graph



Fig. 10 CLSTM model summary

Model: "sequential"		
Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 50, 300)	10295100
conv1d (Conv1D)	(None, 48, 64)	57664
conv1d_1 (Conv1D)	(None, 46, 64)	12352
batch_normalization (Batch Normalization)	(None, 46, 64)	256
max_pooling1d (MaxPooling1D)	(None, 23, 64)	0
conv1d_2 (Conv1D)	(None, 21, 128)	24704
conv1d_3 (Conv1D)	(None, 19, 128)	49280
batch_normalization_1 (Batch Normalization)	(None, 19, 128)	512
max_pooling1d_1 (MaxPooling1D)	(None, 9, 128)	0
bidirectional (Bidirectional)	(None, 9, 200)	183200
bidirectional_1 (Bidirectional)	(None, 9, 200)	240800
flatten (Flatten)	(None, 1800)	0
dropout (Dropout)	(None, 1800)	0
dense (Dense)	(None, 256)	461056
dense_1 (Dense)	(None, 2)	514
=====		
Total params: 11325438 (43.20 MB)		
Trainable params: 11325054 (43.20 MB)		
Non-trainable params: 384 (1.50 KB)		
=====		

complex relationships within textual data, making it well-suited for advanced text analysis tasks.

7187, with a mere 65 misclassifications. This results in an impressive accuracy level of 96%.

Confusion Matrix-CLSTM

Figure 11 displays the confusion matrix for the CLSTM model. A detailed examination of the matrix showcases remarkable performance. Out of the total 14,504 samples, the model adeptly discerns 6738 as fake and erroneously categorizes 514. Likewise, for real targets, it accurately predicts

Accuracy & Loss Plot Graph for CLSTM

An accuracy plot graph visually represents the performance of a model over a range of iterations or epochs during the training process. It plots the accuracy score on the y-axis against the number of iterations or epochs on the x-axis. As training progresses, the plot shows how well the model is

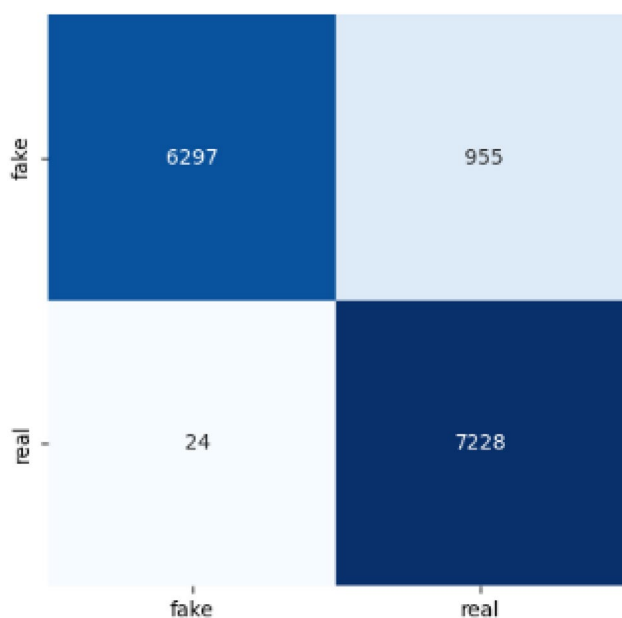


Fig. 11 Confusion matrix for CLSTM

learning from the data. Initially, the accuracy may fluctuate, indicating the model's struggle to understand complex patterns. However, with more iterations, the accuracy typically improves, signifying better generalization to unseen data. The plot may exhibit a plateau, suggesting diminishing returns from further training. Monitoring an accuracy plot helps practitioners make informed decisions about when to stop training to prevent overfitting. Additionally, it provides insights into the model's learning curve, aiding in fine-tuning and optimizing its performance for real-world applications (Fig. 12).

Loss Plot Graphs

A loss plot graph visually represents the performance of a machine learning model during training. It displays the value of the loss function on the y-axis and the number of training iterations or epochs on the x-axis. Initially, the loss is typically high, indicating that the model's predictions are far from

Fig. 12 Accuracy plot graph

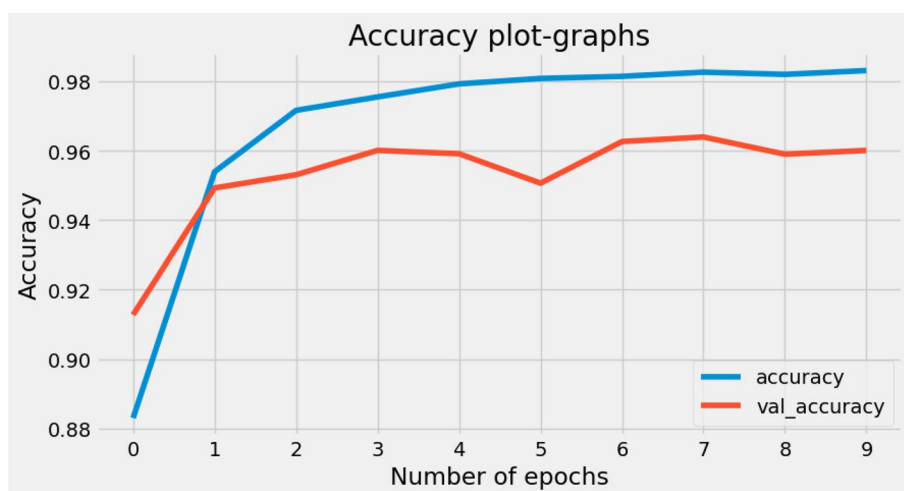
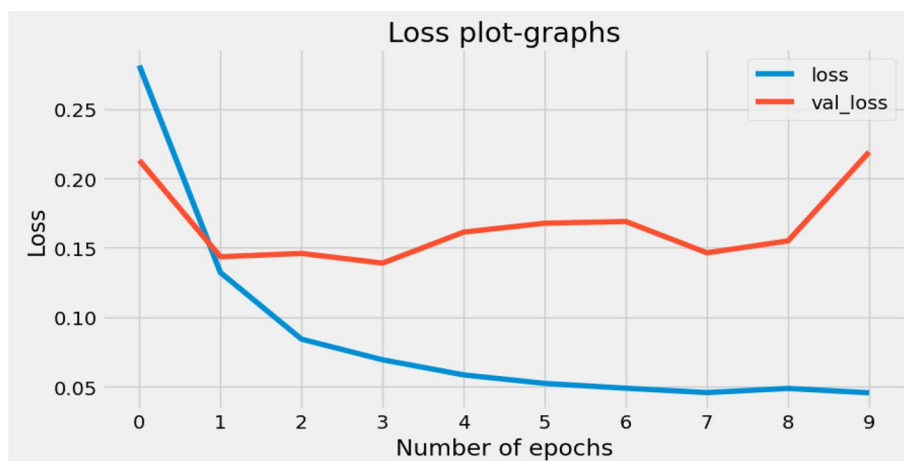


Fig. 13 Loss plot graph



the actual values. As training progresses, the loss gradually decreases, signifying improved accuracy. The graph should show a steady decline, although there may be fluctuations. If the loss suddenly spikes, it could indicate overfitting, where the model is fitting too closely to the training data and may perform poorly on new, unseen data. Conversely, a loss that stagnates may imply that the model has reached its learning capacity. A well-constructed loss plot graph is crucial for understanding the training process and making informed decisions about model adjustments (Fig. 13).

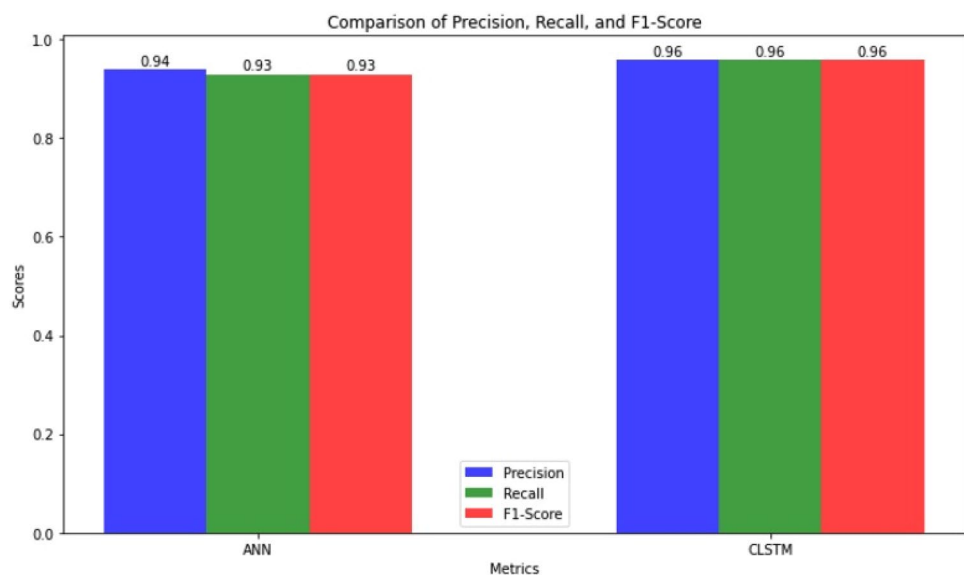
Comparison of CLSTM with ANN

The table provides a comparative summary of classification performance metrics for two models: ANN (artificial neural network) and CLSTM (convolutional long short-term memory). For the ANN model, it exhibits excellent precision of 1.00 for classifying fake news, indicating a very low rate of false positives. However, its recall is 0.87, suggesting it misses around 13% of actual fake news instances. The F1-Score, a balance of precision and recall, is 0.93, signifying robust overall performance. The ANN model achieves

Table 1 Comparison of models

Class	Precision	Recall	F1-Score	Accuracy
<i>ANN</i>				
Fake	1.00	0.87	0.93	0.93
Real	0.88	1.00	0.94	
<i>CLSTM</i>				
Fake	0.99	0.93	0.96	0.96
Real	0.93	0.99	0.96	

Fig. 14 Accuracy comparison graph



an accuracy of 93%. Conversely, the CLSTM model excels in both precision and recall for fake news, with values of 0.99 and 0.93 respectively. This indicates high accuracy in correctly identifying fake news while minimizing false negatives. The F1-Score for fake news is 0.96. For real news, the CLSTM model demonstrates a slightly lower precision of 0.93 but a high recall of 0.99. The overall F1-Score for CLSTM is also 0.96, indicating consistent and reliable performance across both classes. The CLSTM model achieves a commendable accuracy rate of 96%, showcasing its effectiveness in classifying news articles (Table 1).

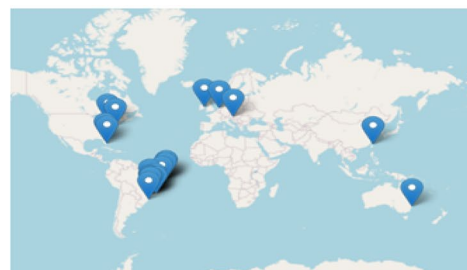
Comparison Graph for ANN and CLSTM

The accuracy comparison between artificial neural networks (ANN) and convolutional long short-term memory (CLSTM) models reveals interesting insights into their respective performance. In this evaluation, both models were assessed on a classification task, likely involving text data. This suggests its proficiency in learning complex patterns within the data. On the other hand, this could be attributed to CLSTM's ability to simultaneously capture spatial features through convolutional layers and sequential dependencies with long short-term memory units. The CLSTM model's accuracy surpasses that of the ANN, indicating its strength in tasks involving grid-like data, such as images or time series.

The comparison highlights that the CLSTM architecture was particularly well-suited for the given task. It excelled in discerning intricate patterns in the data, potentially due to its combination of convolutional and recurrent layers. However, it is crucial to note that the effectiveness of a model heavily depends on the specific nature of the data and the task

Fig. 15 Geo locations from where fake news were originated

	name	latitude	longitude	lived	visited
0	Olinda	-7.998127	-34.908203	yes	yes
1	Recife	-8.052200	-34.928600	yes	yes
2	Miami	25.761700	-80.191800	no	yes



at hand. Thus, while the CLSTM model outperformed the ANN in this particular scenario, different tasks or datasets might yield different results, emphasizing the importance of selecting the appropriate model for each specific application (Fig. 14).

Geo-Location

In the implementation phase, the geographical location of a user's tweets is established by extracting longitude and latitude information from the dataset using their user IDs. Additionally, the tweets of various individuals are scrutinized to determine both their geographical origin and the timestamp of the data, derived from their log reports. Consequently, these models effectively discern the distinctive nature of tweets, categorizing them as positive or negative feedback, regular discourse, hate speech, or commendable or unfavorable comments. This model distinguishes itself from previous approaches by incorporating the monitoring of both geo-location and temporal aspects. Upon the immediate posting of a tweet, the model conducts a thorough analysis, accurately categorizing the content as either conventional news or potentially containing fake messages. Figure 15 illustrates the locations of fake news on a geo-map.

Conclusion

The hybrid deep learning model comprising both CNN and LSTM components has demonstrated promising results in the task of fake news detection in social media. The ANN model achieved impressive precision and recall rates for both fake and real news, with an overall F1-Score of 0.93 and an accuracy of 93%. On the other hand, the CLSTM model excelled in terms of precision and recall for both classes, yielding an overall F1-Score of 0.96 with accuracy of 96%. This suggests that the CLSTM model outperforms the ANN model in terms of overall accuracy and reliability. However, further analysis and testing would be valuable to confirm the robustness of these results and to consider other factors like computational efficiency and scalability in real-world applications. Overall, this hybrid

deep learning approach represents a promising tool in the ongoing battle against fake news and misinformation, offering the potential to enhance the trustworthiness of information disseminated through social media platforms.

It is crucial to acknowledge the added value of incorporating geo-map and geo-tag information in the detection process. By considering geographical aspects, the model could further enhance its ability to discern the authenticity of information circulating on social media. Integrating geo-location data provides an additional layer of context, enabling the model to identify regional patterns and potentially distinguish between local and global instances of fake news.

Future Work

Future work for this hybrid deep learning model includes fine-tuning and optimization to enhance its performance, especially in handling different types of misinformation. Additionally, expanding the model to process multimodal data, incorporating images and videos, could provide a more comprehensive solution. Real-time monitoring capabilities should be developed to swiftly identify and counter fake news as it spreads. These avenues of improvement and innovation will be essential in the ongoing fight against fake news, ensuring the continued integrity of information in an ever-evolving digital landscape.

Author Contribution Not applicable.

Funding Not applicable.

Data Availability Not applicable.

Declarations

Conflict of interest Not applicable.

Research Involving Human and/or Animals Not applicable.

Informed Consent Not applicable.

References

1. Hanshal OA, Ucan ON, Sanjalawe YK. Hybrid deep learning model for automatic fake news detection. *Appl Nanosci*. 2023;13(4):2957–67.
2. Jing J, et al. Multimodal fake news detection via progressive fusion networks. *Inf Process Manag*. 2023;60(1):103120.
3. Rawat G, et al. Fake news detection using machine learning. In: 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), IEEE. 2023.
4. Lai CM, Chen MH, Kristiani E, Verma VK, Yang CT. Fake news classification based on content level features. *Appl Sci*. 2022;12(3):1116.
5. Mridha MF, Keya AJ, Hamid MA, Monowar MM, Rahman MS. A comprehensive review on fake news detection with deep learning. *IEEE access*. 2021.
6. Nasir JA, Khan OS, Varlamis I. Fake news detection: a hybrid CNN-RNN based deep learning approach. *Int J Inf Manag Data Insights*. 2021;1(1): 100007.
7. Khan JY, Khondaker MTI, Afroz S, Uddin G, Iqbal A. A benchmark study of machine learning models for online fake news detection. *Mach Learn Appl*. 2021;4: 100032.
8. Kaliyar RK, Goswami A, Narang P. A hybrid model for effective fake news detection with a novel COVID-19 dataset. In: ICAART. 2021;2:1066–1072.
9. Zeng J, Zhang Y, Ma X. Fake news detection for epidemic emergencies via deep correlations between text and images. *Sustain Cities Soc*. 2021;66: 102652.
10. Aslam N, Ullah Khan I, Alotaibi FS, Aldaej LA, Aldubaikil AK. Fake detect: a deep learning ensemble model for fake news detection. *Complexity*. 2021;2021:1.
11. Ashik SS, Apu AR, Marjana NJ, Islam MS, Hasan MA. M82B at CheckThat! 2021: Multiclass fake news detection using BiLSTM. 2021.
12. Chauhan T, Palivela H. Optimization and improvement of fake news detection using deep learning approaches for societal benefit. *Int J Inf Manag Data Insights*. 2021;1(2): 100051.
13. Choudhary M, Chouhan SS, Pilli ES, Vipparthi SK. BerConvoNet: a deep learning framework for fake news classification. *Appl Soft Comput*. 2021;110: 107614.
14. Sadr MM. The use of LSTM neural network to detect fake news on Persian twitter. *Turkish J Comput Math Edu (TURCOMAT)*. 2021;12(11):6658–68.
15. Seyam A, Bou Nassif A, Abu Talib M, Nasir Q, Al Blooshi B. Deep learning models to detect online false information: a systematic literature review. In: The 7th Annual International Conference on Arab Women in Computing in Conjunction with the 2nd Forum of Women in Research, 2021;1–5.
16. Chitra NT, Anusha R, Kumar SH, Chandana DS, Harika C, Kumar VU. Satellite imagery for deforestation prediction using deep learning. In: 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, 2021;522–525.
17. Islam MR, Liu S, Wang X, Xu G. Deep learning for misinformation detection on online social networks: a survey and new perspectives. *Soc Netw Anal Min*. 2020;10(1):1–20.
18. Umer M, Imtiaz Z, Ullah S, Mehmood A, Choi GS, On BW. Fake news stance detection using deep learning architecture (CNN-LSTM). *IEEE Access*. 2020;8:156695–706.
19. Kaliyar RK, Goswami A, Narang P, Sinha S. FNDNet—a deep convolutional neural network for fake news detection. *Cogn Syst Res*. 2020;61:32–44.
20. Aniyath A. A survey on fake news detection by the data mining perspective. *Int J Inf Comput Sci*. 2019;6(1):9–28.
21. Zhang J, Dong B, Philip SY. Fakedetector: effective fake news detection with deep diffusive neural network. In: 2020 IEEE 36th International Conference on Data Engineering (ICDE), IEEE, 2020;1826–1829.
22. Lee DH, Kim YR, Kim HJ, Park SM, Yang YJ. Fake news detection using deep learning. *J Inf Process Syst*. 2019;15(5):1119–30.
23. Drif A, Hamida ZF, Giordano S. Fake news detection method based on text-features. In: The Ninth International Conference on Advances in Information Mining and Management (IMMM), 2019.
24. Ramya SP, Eswari R. Attention-based deep learning models for detection of fake news in social networks. *Int J Cogn Inf Nat Int (IJCINI)*. 2021;15(4):1–25.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.