Multimedia Tools and Applications (2022) 81:5587–5620
https://doi.org/10.1007/s11042-021-11782-3

CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT

Balasubramanian Palani[1] · **Sivasankar Elango**[1] · **Vignesh Viswanathan K**[2]

Received: 30 April 2021 / Revised: 20 August 2021 / Accepted: 25 November 2021 /
Published online: 28 December 2021

Abstract
The progressive growth of today's digital world has made news spread exponentially faster on social media platforms like Twitter, Facebook, and Weibo. Unverified news is often disseminated in the form of multimedia content like text, picture, audio, or video. The dissemination of such false news deceives the public and leads to protests and creates troubles for the public and the government. Hence, it is essential to verify the authenticity of the news at an early stage before sharing it with the public. Earlier fake news detection (FND) approaches combined textual and visual features, but the semantic correlations between words were not addressed and many informative visual features were lost. To address this issue, an automated fake news detection system is proposed, which fuses textual and visual features to create a multimodal feature vector with high information content. The proposed work incorporates the bidirectional encoder representations from transformers (BERT) model to extract the textual features, which preserves the semantic relationships between words. Unlike the convolutional neural network (CNN), the proposed capsule neural network (CapsNet) model captures the most informative visual features from an image. These features are combined to obtain a richer data representation that helps to determine whether the news is fake or real. We investigated the performance of our model against different baselines using two publicly accessible datasets, Politifact and Gossipcop. Our proposed model achieves significantly better classification accuracy of 93% and 92% for the Politifact and Gossipcop datasets, respectively, compared to 84.6% and 85.6% for the SpotFake+ model.

Keywords Fake news detection · Deep learning · BERT · Capsule neural network · Routing-by-agreement

✉ Balasubramanian Palani
balaiiits@gmail.com

Extended author information available on the last page of the article

1 Introduction

The tremendous growth in information and communication technology (ICT) and high-speed internet, people worldwide are more interested in reading the news about the events on social networking platforms like Twitter[1], Facebook[2], and Weibo[3]. Misinformation creators were intentionally flooding falsified and unverified information for various political and commercial purposes. For example, during the 2016 presidential election campaign in the United States [3], false news about Donald Trump, Hillary Clinton, and their political parties were disseminated. The twitter account of Associated Press (AP) was hacked and a post titled "Two Explosions in the White House and Barack Obama is injured" was published [35]. As a result of this rumor, the stock market lost 130 billion dollars in a matter of minutes Recently, there has been a surge of false information on social media about the COVID-19 (Corona virus disease 2019) disease [21]. Due to misconceptions of information about COVID-19, people got confused about the nature of the disease, its causes, and its preventive measures. Thus, the detection of fake news is essential at the early stage, which will help the society and the government to come out of the negative influences.

In general, fake news detection (FND) methods can be categorized into two types: *social context-based* and *content-based* methods [6, 12, 39, 60]. The former is more concerned with user engagement data such as comments, reposts, and ratings, while the latter is associated with the article's news content (title, text, image, and video). The *social context-based* methods can also be divided into two categories: *propagation structure-based* and *post-based* methods [58]. The *propagation structure-based* methods concentrate on propagation patterns or trends of fake news on social networks, while *post-based* methods examines the opinions or emotions expressed by users in their posts [13, 23, 24, 26, 27, 38, 50, 51]. Due to the unstructured nature of the data, these two types of social-context techniques face the following challenges: data collection and analysis, noisy data and missing data. Hence, the focus of this research is on a *content-based* strategy. The content-based methods are more straightforward and convenient way to detect fake news, particularly at an early stage.

Unimodal content-based FND approaches are ineffective at detecting fake news because they use textual [1, 2, 5, 7, 9, 18, 22, 25, 30, 31, 33, 41, 47, 54, 55] and visual features [4, 11, 14, 28, 34, 56, 59] separately. In general, no underlying semantics were discovered among these features, resulting in poor classification. As a result, the multimodal content-based FND approaches have been designed to recognize fake news by integrating textual and visual features in order to increase the effectiveness of the model.

Existing multimodal FND solutions have failed to improve the performance of the FND problem because they are unable to generate enhanced feature representation from the visual information of the news report. A convolution neural network (CNN) model called VGG-19 (Visual Geometry Group) has been used to capture visual features in several studies [20, 43, 44, 49, 53]. The challenges of CNN model are as follows: i) It requires lot of training data to improve the generalization of the model ii) It takes more training time iii) Due to the pooling operation, important information is lost. Hence, CNN drops most important features from an image. Such less informative visual features combined with textual features have not resulted

[1] https://twitter.com/
[2] https://www.facebook.com/
[3] https://weibo.com/

in a richer data representation that allows for the identification of fake news in multimodal news articles. Therefore, the CapsNet model is proposed to solve this problem by capturing highly informative visual features from an image. Also, it takes lesser time compared to CNN for training the data. Furthermore, a pre-trained language model named BERT has been used for textual feature extraction, which is more efficient than the other word embedding techniques and sequence-to-sequence models [8]. It follows the encoder module of the transformer architecture and reads the entire input sequence at once in both directions (left-to-right and right-to-left) to capture contextual relationships among words in a sentence.

In this paper, a new model named CB-Fake is proposed to improve the performance of fake news detection. In the name CB-Fake, C refers to the CapsNet model, B refers to the BERT model and the word Fake refers to the fake news detection. An end-to-end framework is developed, that combines CapsNet and BERT models for fake news detection. The steps involved in the CB-Fake model are as follows: First, the news articles are preprocessed and converted to the vector representation. Then the textual features from the news content have been obtained using BERT. It exploits a self-attention mechanism of transformer architecture, and hence it efficiently extracts the underlying semantic relationships of words in a sentence [45]. Furthermore, the proposed work's key contribution is the use of the capsule neural network model, which is the first attempt to extract informative visual features from the images of news articles using the routing-by-agreement algorithm [16]. Finally, a richer data representation is obtained by combining high-level textual and visual features, which resulted in classification accuracy of 93% for politifact and 92% for gossipcop, when compared to other state-of-the-art approaches in FND. The fully connected layer with softmax activation has been used to recognize whether the news is fake and real. The main contributions of this work are summarized as follows:

- To the best of our knowledge, this is the first multimodal automated fake news detection work based on BERT and CapsNet.
- The proposed CB-Fake model uses textual and visual information from social media news articles to determine if the news is fake or real.
- The proposed model improves classification performance by concatenating the textual and the visual features. The former is extracted using BERT and the latter is captured by CapsNet.
- Experiments are carried out on two publicly accessible datasets from the real world. The results show that our proposed CB-Fake model outperforms state-of-the-art multimodal FND models in detecting fake news.

This paper is structured as follows: Section 2 presents a brief overview of the related works of fake news identification task. In Section 3, the proposed CB-Fake model is described in detail. Experimental setup and the evaluation metrics for the experiments are discussed in Section 4. In Section 5, the experimental results and performance analysis of the proposed model has been discussed and the article is concluded with the possibilities of future work directions in Section 6.

2 Related works

A brief overview of notable relevant works to the proposed model is given in this section. According to the literature, a traditional Machine Learning (ML) [1, 2, 9, 22, 30, 31, 33] and Deep Learning (DL) [5, 7, 18, 25, 41, 47, 54, 55] techniques effectively used textual, visual, and social-context features to solve the automated fake news detection problem. Furthermore, existing research is divided into two categories: unimodal and multimodal fake news identification. To detect fake news, the former uses either one of the content-based [1, 2, 5, 7, 9, 18, 22, 25, 30, 31, 33, 41, 47, 54, 55] or social-context-based features [13, 23, 24, 26, 27, 38, 50, 51], while the latter uses a combination of any single modality feature [20, 43, 44, 49, 53].

2.1 Unimodal fake news detection methods

Researchers used text, visual, and social context-based features in unimodal FND methods to verify the genuineness of a news article, which are explained in the following subsections.

2.1.1 Social context based fake news detection

Social context features represent the active interaction of users by analysing their posts, comments, tags, ratings, etc. about the emerging news on social media. Table 1 summarizes the existing social-context based fake news detection models. Wu et al. [50] used LSTM network over propagation paths and derived embeddings of user-profile information on social media to identify fake news. Ma et al. [27] designed a recursive neural model, which exploits tree structures neural network to learn representation of each tweet. Liu el al. [23] built a time series classifier model based on RNN and CNN to detect fake news through the propagation path of classified news. Guo et al. [13] investigated HSA-BLSTM model that extracts textual and social-context features for fake news detection. Ma et al. [26] and Li et al. [24] designed a model that accepts the user stance in multi-task learning to detect the rumors efficiently. Ke Wu et al. [51] captured the propagation patterns of the news in the form of graphs. Savyan and Bhanu [38] built the UbCadet model, which uses unsupervised learning to detect compromised accounts on the twitter platform. However, social context-based features are very noisy, unstructured in nature and labor-intensive to collect when a new event emerges on social media platforms. Thus, the proposed method solely relies on the news article's content to effectively judge the veracity of the news.

2.1.2 Textual features based fake news detection

Works based on textual content extracts the meaningful linguistic features such as statistical, lexical, syntactic, semantic and style based features from the posts of news articles to verify the falsification of news [6]. Table 2 summarizes the existing textual-based fake news detection models.

Machine learning based methods: Ozbay and Alatas [31] proposed a two-step approach for fake news identification on text information; the former performs the pre-processing steps and the textual feature vector is then applied to twenty three intelligent supervised classifiers for experimental evaluation later. Faustini Covoes [9] designed a language and platform-independent model for fake news detection by extracting the text features using

Table 1 A summary and comparative study of existing social-context based fake news detection

Work	Model	Dataset	Description	Limitations
Wu et al. [50]	LSTM, RNN	Twitter	Focused on diffusion network information, identified propagation pathways of social media messages, solved data sparsity problem	Domain knowledge expertise required, Content of the news has not been used.
Ma et al. [27]	Recursive NN	Twitter-15, Twitter-16	Used propagation tree to learn the representations from the structural and textual properties	Difficult in predicting non-rumors, Not used user information features
Liu el al. [23]	RNN, CNN	Weibo, Twitter 15, Twitter 16	Captured both local and global variations of user characteristics along propagation paths	User characteristics to identify the users' tendency has not been analysed
Guo et al. [13]	HSA-BLSTM	Weibo, Twitter	Learned most useful information and combined with social context features	Less accuracy
Ma et al. [26]	RNN, GRU	LIU, PHEME, FNC	Unified approach for multi-task learning such as rumour detection and stance classification, task-invariant and task-specific features are learned	Users trustworthiness evaluation was not incorporated
Li et al. [24]	LSTM, attention	RumorEval, PHEME	In rumor detection layer, user credibility information has been incorporated. In addition, attention mechanism is introduced in the rumor detection task.	Accuracy is very less
Ke Wu et al. [51]	Hybrid SVM	Sina Weibo	Extracted propagation patterns in the form of graphs and then hybrid SVM is used for classification. Random walk graph kernel has been used to model the propagation tree	Deep learning models are not explored.
Savyan et al. [38]	UbCadet model (k-NN, ensemble)	Twitter, Yelp	Captured user-behavioural characteristics from the tweet text content, hashtag, post time and geolocation	Semantic analysis has not been considered on tweet contents.

Table 2 A summary and comparative study of existing textual-based fake news detection

Work	Model	Dataset	Description	Limitations
Ozbay and Alatas [31]	TF-IDF, ML models	ISOT	Extracted textual feature vector using TF-IDF. Twenty-three supervised classifiers are used	Different word embedding techniques, ensemble and DL based classifiers were not used
Faustini Covoes [9]	BoW, word2vec, RF, SVM	FakeBrCorpus, TwitterBR, btvlife-style	Extracted the textual features using word embedding techniques	DL models has not been utilized
Ozbay and Alatas [30]	GWO, SSO	BuzzFeed, Liar	A meta-heuristic algorithm was used to preserve the global search ability	Word embedding techniques and hybrid model were not used
Perez-Rosas et al. [33]	Linear-SVM	FakeNewsAMT, Celebrity news dataset	Used linguistic-based features such as lexical, syntactic, and semantic level features, Cross-domain classification also performed	DL models were not employed
Ahmed et al. [1]	TF-IDF, Linear-SVM	ISOT	The feature extractor TF-IDF combined with a Linear-SVM classifier achieved the best performance	DL based methods has not been utilized
Kumar et al. [22]	PSO, ML classifiers	Twitter	Selected an optimal feature set using PSO	Biased to English text only tweets, PSO results has not been compared with other optimization algorithms
Akyol et al. [2]	GBT,MLP, RF	Facebook, Google+, LinkedIn	Obtained datasets in four categories: Microsoft, Economic, Palestine and Obama	Recent DL models and word embedding methods were not used
Ma et al. [25]	RNN, GRU, LSTM	Twitter, Weibo	Feature vector of words in the post extracted using TF_IDF	Different word embedding techniques and hybrid model were not tried
Kaliyar et al. [18]	BERT, CNN	Fakenews (2016 U.S presidential election)	Preserved semantic and long-term dependency in sentences and eliminates the ambiguity issue	Hybrid features and different echo-chambers have not been explored
Asghar et al. [5]	Bi-LSTM, CNN	PHEME	Explored in both directions of the sentence to capture contextual information	The model works on English text datasets and textual features only
Shu et al. [41]	GRU encoder, co-attention	FakeNewsNet	Co-attention mechanism was utilized to discover top-K important sentences and user reviews	Fake-checking contents and user related information have not been utilized

Table 2 (continued)

Work	Model	Dataset	Description	Limitations
Chen et al. [7]	RNN, soft attention	Twitter, Weibo	Collected distinct linguistic features over time. Learned latent representation from paragraph vector	Propagation pattern of rumours were not utilized
Yu et al. [55]	CNN	Twitter, Weibo	Extracted key features from the text and high-level interactions among those features	Prediction accuracy was less and no word embedding methods were used
Wang [47]	CNN, Bi-LSTM	LIAR	Created a larger dataset, Used CNN for the textual-feature extraction and Bi-LSTM for meta-data feature extraction	Obtained less prediction accuracy
Yin et al. [54]	PCA, CNN, SVM	Private dataset	Extracted feature vectors using PCA and CNN	Prediction accuracy was less

different word embedding techniques on five different datasets of three different languages from various social media platforms. Recently, Guo et al. [12] investigated and presented a comprehensive review on recent research challenges, novel techniques, and details of datasets in the field of fake information detection. Ozbay and Alatas [30] utilized a meta-heuristic technique, grey wolf optimizer (GWO) and salp swarm optimizer (SSO) for fake news identification. Perez-Rosas et al. [33] developed two novel datasets for automatic fake news detection, which consists of seven various domains of news. Ahmed et al. [1] introduced a new dataset collected from real-world sources, called ISOT, for fake news identification and employed n-gram analysis with Term frequency - Inverse document frequency (TF-IDF) for feature vector representation to classify the fake news article. Kumar et al. [22] used particle swarm optimization (PSO) for selecting an optimal feature set from the textual content of the tweets, to solve the rumour veracity classification task efficiently. However, these methods are time-consuming and labor-intensive, since it needs hand-crafted features. Akyol et al. [2] designed gradient boost tree (GBT), multilayer perceptron and random forest (RF) to identify fake news).

Deep learning based methods: Nowadays, many researchers adopt into deep learning models, since it has the ability to extract the high-level features automatically from the news contents, and hence it plays a vital role to identify fake news effectively [5, 7, 18, 25, 41, 47, 54, 55]. Ma et al. [25] were the first to introduce Recurrent Neural Network (RNN) to model the textual data sequence for detecting rumours over time. Kaliyar et al. [18] designed a FakeBERT model, which is a combination of BERT and Convolutional Neural Network (CNN) to handle the textual contents in a bidirectional way. Asghar et al. [5] proposed a deep learning model which combines Bidirectional Long Short-Term Memory (Bi-LSTM) and CNN for rumour detection on text data. Co-attention mechanism was utilized by Shu et al. [41] to discover top K important sentences from the articles and top K important user reviews to classify fake news. Chen et al. [7] presented attention-based RNN to collect distinct linguistic features over time for early rumour detection. Yu et al. [55] proposed a CNN model, to identify misinformation. Wang [47] introduced a new dataset, called LIAR dataset, proposed a hybrid deep learning model in which CNN for the textual-feature extraction and Bi-LSTM for meta-data feature extraction is used. Yin et al. [54] employed principal component analysis (PCA) and CNN for feature extraction from the news contents. Although these DL models have been widely applied to the textual content of the news, they have failed to achieve the following: preserving long-term word dependencies, parallelization mechanism in training, and taking input sentences bidirectionally. To address this issue, the pre-trained BERT model was recently introduced to capture high-level context-based textual features from news articles. Hence, the BERT model has been used in this proposed work to extract semantically meaningful textual features from the news articles.

2.1.3 Visual features based fake news detection

Recent studies [34, 59] have proven that the visual features plays a vital role in detecting fake news on multimedia contents. Works [4, 14] dealt with basic features extraction from the attached images. Also the features are still hand-crafted. Hence, it is difficult to represent the complex distributions of visual contents. Zeng et al. [56] explored the forged fake images problem and detects such behaviors by using image splicing techniques. In the work of [11, 28], the authors proposed Generative adversarial networks (GAN) model to recognize forged images. Zhou et al. [59] developed a faster R-CNN model, which takes RGB

stream and noise stream to discover tampering features for image manipulation detection. Qi et al. [34] proposed a framework including three modules: frequency domain module, pixel domain module and fusion module, which learns visual representations to identify fake images. Furthermore, the studies mentioned above were using a CNN model to learn visual features from the image. The main drawback of CNN is its pooling operation, which is either max-pooling or average-pooling. Pooling causes the most informative features to be lost when extracting visual features from an image associated with a news article. To solve this issue, the CapsNet model is proposed in this work to capture most informative visual features. CapsNet has shown a magnificent performance in diverse areas, especially in image recognition and NLP. In addition, CapsNet's popularity attracts researchers to work on real-world problems like machine translation, drug discovery, handwritten text recognition, self-driving cars, healthcare, and emotion detection [32].

In recent years, the majority of news articles have included both text and images. As a result of only using single modality features, the aforementioned unimodal FND approaches cannot distinguish effectively between fake and true news. It is also inadequate and easily gets affected by different surrounding factors. As a consequence, multimodal features can be considered for a better classification of fake or real news. Specifically, fusing textual and visual features plays a vital role in obtaining an enhanced feature representation of news articles.

2.2 Multimodal fake news detection methods

Deep neural networks have been widely used for different multimodal data dependent tasks such as visual question answering [4], image captioning [19] and fake news detection [20, 43, 44, 49, 53]. Table 3 summarizes the existing multimodal fake news detection models. Jin et al. [17] built an attention based RNN model which mines the textual, visual and social context features, and combines them by using attention mechanism. Singh et al. [42] designed an extreme learning machine (ELM) model for various internet-of-things (IoT) applications. Yang.K et al. [52] analysed text and images and then derived user interested tags using adaptive tag (AT) algorithm. Yang et al. [53] proposed TI-CNN (Text Image - CNN) model to capture explicit and hidden features from text and images for fake news detection. Wang et al. [49], proposed an end-to-end framework for fake news detection and event discriminator and named it as Event Adversarial Neural Network (EANN). In multimodal feature extractor part, textual and visual features were extracted using Text-CNN and VGG-19 model respectively. However, this model does not have any clear idea to discover correlations across the modalities. Khattar et al. [20], addressed this issue and built a similar architecture, termed as Multivariational Autoencoder for Fake news detection (MVAE). The primary task of MVAE model is to learn the shared representation or latent vector of multimodal (textual+visual) information from an encoder module. Decoder uses this latent vector for the reconstruction of original samples.

Shivangi et al. [44], built a multimodal framework for fake news detection termed as SpotFake model. This model eliminates the additional task of EANN and MVAE; and also achieved higher accuracy gain for detecting fake news. Compared to the earlier works [20, 49], SpotFake provides a reasonable accuracy gain over EANN and MVAE since it uses the BERT model and pre-trained CNN model on Imagenet database (VGG-19) for textual and visual feature representation respectively. Shivangi et al. [43], designed a framework called SpotFake+, an advanced version of SpotFake [44]. This proposed architecture has the benefit of handling a dataset that consists of full length articles. This model shown

Table 3 A summary and comparative study of existing multimodal fake news detection

Work	Model	Dataset	Description	Limitations
Jin et al. [17]	RNN-attention, LSTM, VGG-19	Twitter, Weibo	Extracts the textual, visual and social-context features and fused all these features by attention mechanism	Obtained very less prediction accuracy
Singh et al. [42]	PCA, K-means, ELM	NSL-KDD	Data pre-processing using PCA and K-means. ELM model is adopted to the maximum number of IoT applications	Different feature extraction methods and DL based models were not utilized
Yang.K et al. [52]	Adaptive tag (AT)	Toutiao news	Extracted new tags from the images and texts. Based on user's feedback, AT algorithm selects the user interested tags	DL model has not been used
Yang et al. [53]	TI-CNN	U.S president election news (Kaggle)	TI-CNN model was used to capture explicit and hidden features from text and images for fake news detection	User's characteristics and social network structures were not used
Wang et al. [49]	EANN (Text-CNN, VGG-19)	Twitter, Weibo	Obtained event-invariant features by the event discriminator component of adversarial network	Prediction is an additional task and does not have any clear idea to discover correlations across the modalities
Khattar et al. [20]	MVAE (Encoder-Decoder)	Twitter, Weibo	Learned the shared representation or latent vector of multimodal information. Predicted the fake news based on the latent vector	Fake news prediction is secondary task
Shivangi et al. [44]	SpotFake (BERT, VGG-19)	Twitter, Weibo	Extracted semantically meaningful textual features and visual feature using BERT and VGG-19 respectively	CNN takes more training time and requires huge data collection. Not able to handle full length articles
Shivangi et al. [43]	SpotFake+ (XL-Net, VGG-19)	FakeNewsNet (Politifact, Gossipcop)	Captured the textual (Pre-trained XL-Net) and visual (VGG-19) features	More training time, VGG-19 does not capture most important visual features since it has pooling layer which leads to information loss

improvement over other works [20, 44, 49], since it uses transfer learning to capture the textual (Pre-trained XL-Net) and visual (VGG-19) features within a news article .

In summary, the following are the challenges of existing multimodal FND approaches, and is shown in Table 4. Although there are various sequence models (RNN, LSTM, Bi-LSTM, etc.) and the Text-CNN model for textual content processing, these models are deficient in learning long-term dependencies between words, lack of parallelization in training and sequential access to the input sentence. Hence, the BERT model is used in this study to solve aforementioned issue. BERT model uses the encoder module of the transformer architecture to detect the existence of high-level contextual word embeddings with the help of its self-attention function, parallelization of training, and bi-directionality of input sentence handling. Furthermore, previous research works used CNN to extract visual features, but it cannot retrieve more informative features due to its pooling operation and translational invariance property. To deal with the problem of information extraction in CNN, CapsNet has been introduced. The most important visual features from the picture of the news articles are discovered using the Routing-by-agreement algorithm and the Margin loss function. Furthermore, to improve the performance of FND, the proposed CB-Fake model fuses semantically meaningful textual features with the informative visual features to obtain an enhanced feature vector representation for the given news article. The enhanced feature vector is flattened and then it is passed to the classification layer. The simple feed forward neural network (FFN) with softmax activation function has been used in the classification layer to predict whether the news article is fake or real based on the probability values. In the following section, the proposed CB-Fake model is described in detail. The abbreviation used in this paper is shown in Table 5.

3 Preliminaries

3.1 Bilinear encoder representations from transformers (BERT)

In recent years, the state-of-the-art pre-trained word encoding language model, BERT [8], received greater attention from a wide range of research communities. It solves the significant hindrances on downstream Natural Language Processing (NLP) tasks such as, question answering, natural language interference, and sentiment analysis by achieving the best performance. Since both RNN and CNN receives a input sequence sequentially, it is challenging to learn long-term dependencies between words in the input sentences. To address this problem, Google AI research team introduced a pre-trained language representation model, named the BERT. It captures underlying semantic and contextual meaning from the input words and sentences by randomly masking word tokens and representing each masked word with a vector. The key components of BERT model are discussed in the following section.

3.1.1 Transformers

In BERT, word encoding is obtained from the raw sentence of the news articles. It is based on the transformer architecture [45], which consists of an encoder-decoder configuration generally used in neural machine translation. *self-attention mechanism* is performed in this architecture, which is responsible for learning the most relevant part of the input sequence. Hence, it captures the long-range dependencies in the word

Table 4 A comparison study of proposed model with existing techniques

Feature	Existing models	Limitations	Proposed solution and its merits
Textual [1, 5, 7, 9, 25, 31, 47, 54, 55]	TF-IDF, BoW, word2vec, RNN, LSTM, Bi-LSTM, CNN, Text-CNN	Fails to extract semantic-based relationship among words. processing of input sequence is either from left to right or right to left. Only one word is taken at a time.	BERT model is used. It is a pre-trained model and it follows the transformer architecture, in which the multi-head attention is used to preserve the semantic relations among words. Masked language modeling (MLM) and Next sentence prediction (NSP) tasks are introduced
Visual [20, 43, 44, 49, 53]	VGG-19 (CNN model)	Taken more training time. Larger dataset is required for better generalization. Due to the pooling operation, it fails to extract informative visual features. Also, it consumes more number of hyperparameters while training the data.	CapsNet model is used. It requires less training data and incurring less training time compared to CNN. Routing-by-agreement algorithm is used, in which squashing activation function is performed. Margin loss function is introduced. Number of hyperparameters are smaller than the CNN

Table 5 Abbreviations used in this paper

Abbreviation	Expansion
BERT	Bidirectional encoder representations from transformers
Bi-LSTM	Bidirectional long short-term memory
CapsNet	Capsule neural network
CB-Fake	CapsNet BERT – Fake
CCL	Class capsule layer
CNN	Convolutional neural network
COL	Convolutional layer
EANN	Event adversarial neural network
FND	Fake news detection
FFN	Feed forward neural network
GAN	Generative adversarial network
GRU	Gated recurrent unit
GWO	Grey wolf optimization
LSTM	Long short-term memory
MLM	Masked language model
MVAE	Multivariational autoencoder
NB	Naive Bayes
NLP	Natural language processing
NSP	Next sentence prediction
PCA	Principal component analysis
PCL	Primary capsule layer
PSO	Particle swarm optimization
RF	Random forest
RNN	Recurrent neural network
SGD	Stochastic gradient descent
SSO	Salp swarm optimization
SVM	Support vector machine
TF-IDF	Term frequency – Inverse document frequency
TI-CNN	Text Image – CNN
VGG-19	Visual Geometry Group – 19

sequence. An encoder block represents a given input sequence in a vector form, and a decoder block takes that encoded vector and generates another sequence. In addition, encoder has divided into two layers: *self-attention layer* and *feed-forward neural network layer*. The transformer model uses a self-attention mechanism, called *"scalar dot-product attention"*, which chooses the most important and relevant part of the input sequence. The input sequence consists of queries and keys of dimension dim_k, and values of dimension dim_v. The dot product is computed between the query to every key, then scaled by square root of dim_k, and a softmax activation function is applied to find the weights on the values. In reality, (1) is used to compute the attention function on a group of queries concurrently, organized together into a matrix Q. Let the matrix K represent the collection of keys, and the matrix V consist of the values. The

three matrices Q, K, and V are learned during the training phase. The attention matrix $Att\_Mat(Q, K, V)$ is calculated in (1) as follows:

$$Att\_Mat(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{dim_k}}\right)V \tag{1}$$

In this transformer architecture, multi-head attention blocks are used to repeat the attention function n-times (n=8), to obtain n-attention matrices, which are then concatenated and multiplied by the matrix, W^O, to get an output. Then, the resulting values are fed into the normalization block of the transformer. The multi-head attention $Multi\_Head(Q, K, V)$ is computed in (2) as follows:

$$Multi\_Head(Q, K, V) = Concat\left(head_1, head_2, \ldots\ldots, head_n\right)W^O \tag{2}$$

where, $head\_i$=$Att\_Mat(QW_i^Q, KW_i^K, VW_i^V)$

3.1.2 BERT-base model

In BERT, two unsupervised tasks such as Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) have been applied for pre-training the model. The BERT model achieved the best results in 11 Natural Language Understanding (NLU) tasks. The BERT model's credibility can be inferred from the fact that Google uses it in its search algorithms and is now rolling out for 70 different languages worldwide [29, 36].

A deep language representation model BERT comprises of two steps: pre-training and fine-tuning. "Pre-training", manipulates two unsupervised based NLP tasks: 1) MLM, which finds the randomly masked input tokens from the input sentence, and 2) NSP, which can be used to identify whether two input sentences are contiguous to each other or not. "Fine-tuning" is the succeeding stage and is represented in Fig. 1. It emphasizes downstream applications; generally, at the top of the last encoder layer, one or additional fully-connected layers are added to the network depending on the application. The primary BERT model appears in two versions: i) BERT-base, ii) BERT-large. Table 6 summarizes the variations of the original BERT model in terms of the number of layers, also called transformer blocks, their hidden layer size, the total number of attention heads, and the total number of parameters used.

3.2 Capsule neural network (CapsNet)

CapsNet, a new sensation in deep learning, was introduced by Geoffrey Hinton [16]. CapsNet has demonstrated superior performance in a number of areas, including image recognition and NLP. CapsNet solves the significant shortcomings of CNNs, such as i) Requiring more training data to generalize a model ii) Inability to identify the position and the pose information of objects in an image, called translational invariance property. To date, a few works explore the achievement of CapsNet [46, 48] for above mentioned problems which tends to use it for fake news detection [10, 32]. The proposed model CB-Fake, aims to emulate this model for discovering visual features from an image of the news article.

A capsule is a collection of neurons that finds out the presence of objects in the data. This group of neurons can be represented as an activation vector, consisting of instantiation

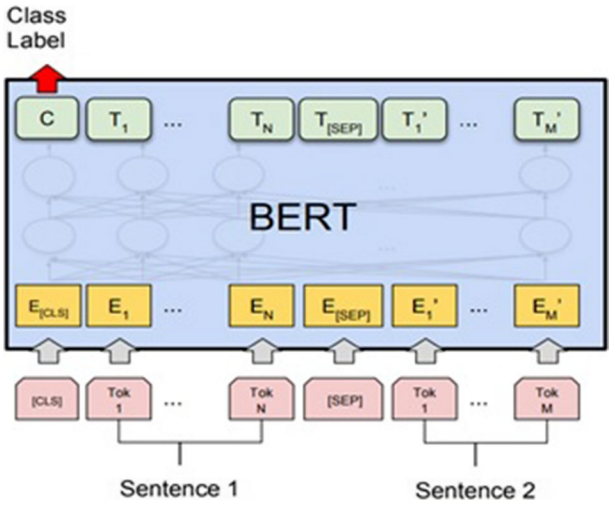


Fig. 1 BERT fine-tuning model [8]

Table 6 Variations of original BERT model

Parameter Name	Value of Parameter	
	BERT-base	BERT-large
Total number of layers	12	24
Hidden layer size	768	1024
Attention heads count	12	16
Total Number of Parameters	110M	340M

parameters or pose parameters of an image or parts of an image. It uses the range of the activity vector to represent the probability that the object exists and its orientations, such as location, posture, and color, to define the instantiation parameters [37]. Capsules vary from conventional artificial neurons by substituting the operation scalar-output detectors of CNNs with vector-output capsules and max-pooling functions, with the routing-by-agreement procedure.

Sabour et al. [15] implemented CapsNet that does not require the pose information of an image as an input. The complete flow diagram of CapsNet is shown in Fig. 2 it comprises of three essential layers: i) convolutional layer, ii) primary capsule layer, and iii) class capsule layer. The role of each layer is explained in the preceding paragraphs.

3.2.1 Convolutional layer - COL

In this layer, the features are extracted based on different kernel sizes and filters from an input image; specifically, the pixel intensities are converted into local feature vectors. Let Z be a set of feature vectors extracted from a given image, say z_1, z_2,....., z_i. It is then sent to the primary capsule layer.

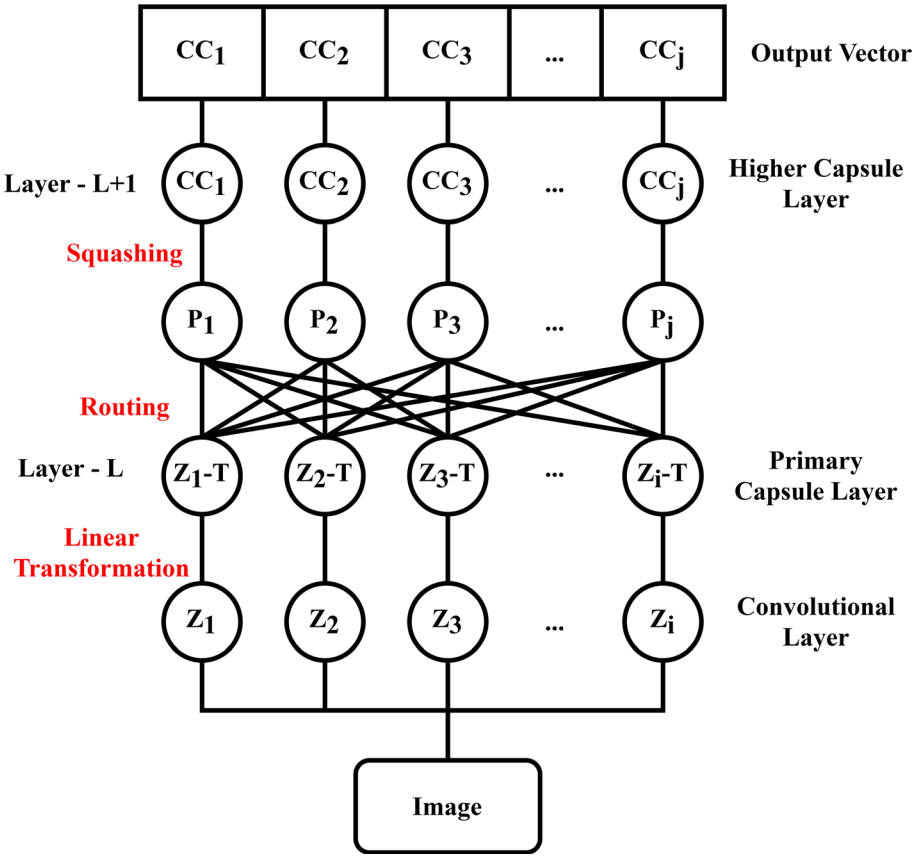


Fig. 2 Complete flow diagram of CapsNet model

3.2.2 Primary capsule layer - PCL

The primary capsule layer's significance is to capture the information about the extracted features and their relationship by using the affine transformation process. PCL is also referred as the lower-level capsule of a network. The network size may grow by adding more capsule layers in the middle if necessary until it reaches the final layer called the classification capsule layer (CCL). The simple FFN with softmax activation function has been used in CCL to recognize the fake news articles.

Let L be the primary capsule layer and each capsule $c_i^{(L)} \in L$ receives $z_i \in \mathbb{R}$ from the previous COL represent an activation vector of an entity in an image. It consists of appropriate instantiation parameters such as pose, color, deformation, hue, etc., for a particular part of an image. The activation vector z_i of the capsule $c_i^{(L)}$ from the layer L is passed to every capsule in the adjacent layer L+1. The capsule $c_j^{(L+1)} \in L+1$ accepts z_i and then computes the linear combination with the weight matrix W_{ij}, which is represented as $\widehat{Z_{j|i}}$ in (3). The computed resulting vector $\widehat{Z_{j|i}}$ represents the transformation of an entity of the capsule $c_i^{(L)}$ at level L described by capsule $c_j^{(L+1)}$ at level L+1.The resulting vectors $\widehat{Z_{j|1}}$, $\widehat{Z_{j|2}}$, ..., $\widehat{Z_{j|i}}$ are known as prediction vectors of primary capsule $c_1^{(L)}$, $c_2^{(L)}$, ... , $c_i^{(L)}$ respectively.

In general, $\widehat{Z_{j|i}}$ indicates the significance of the primary capsule $c_i^{(L)}$ to the next layer capsule $c_j^{(L+1)}$.

$$\widehat{Z_{j|i}} = W_{ij} z_i \tag{3}$$

where, W_{ij} is the translation weight matrix between the capsule $c_i^{(L)} \in L$ and capsule $c_j^{(L+1)} \in L+1$. The capsule was routed from the lower layer to the higher layer using a dynamic routing or routing-by-agreement algorithm. By using this routing algorithm, the agreement between the capsule $c_i^{(L)} \in L$ and capsule $c_j^{(L+1)} \in L+1$ can be computed using (4), and it is represented as P_j.

$$P_j = \sum_{i=1}^{n} \alpha_{ij} \widehat{Z_{j|i}} \tag{4}$$

where, P_j gives a prediction to choose the class capsule $c_j^{(L+1)}$ by the single primary capsule $c_i^{(L)}$ in layer. Two capsules $c_i^{(L)}$ and $c_j^{(L+1)}$ are related to each other if the agreement's value is high; otherwise, they are different. Hence, the value of the coupling coefficient α_{ij} is increased; otherwise, the value of α_{ij} will be reduced. The coupling coefficient α_{ij} is a scalar component between the capsules $c_i^{(L)}$ and $c_j^{(L+1)}$. It can be calculated for all the capsule $c_i^{(L)}$ in the layer L using (5).

$$\alpha_{ij} = \frac{\exp(b_{ij})}{\sum_n \exp(b_{in})} \tag{5}$$

A non-linear activation function, squashing, has been computed on P_j to adjust the length of the output vector between 0 and 1. Equation (6) is utilized to compute the squashing function for all capsule $c_j^{(L+1)} \in L+1$, and it is referred as CC_j.

$$CC_j = \frac{\|P_j\|^2}{1+\|P_j\|^2} \frac{P_j}{\|P_j\|} \tag{6}$$

The logits value or similarity score, b_{ij} will be updated by (7) for all capsule $c_i^{(L)} \in L$ and capsule $c_j^{(L+1)} \in L+1$, till the maximum number of iterations t, rather than until convergence.

$$b_{ij} = b_{ij} + < \widehat{Z_{j|i}}, CC_j > \tag{7}$$

3.2.3 Class capsule layer – CCL

Class capsule layer or higher-level capsule layer is essential to represent the resultant feature vector F_{visual} for predicting class label. The number of capsules is equal to the number of classes in the classification task. The output vector of each capsule in this layer represents the probability of a particular entity of an image is present.

The routing-by-agreement algorithm is illustrated in Algorithm 1. In this algorithm, input vector is passed from the Algorithm 2, which is obtained from $I = \{I_1, I_2, ..., I_m\}$. Line 1 performs linear transformation between weight vector and input vector using (3). Line 3 to Line 7 computes the prediction vector, squashing function and updates

the logits value using (4) to (7) for t iterations. Finally the output visual vector CC_j is returned to the Algorithm 1, which is stored into F_{visual}.

Algorithm 1 Routing_by_agreement algorithm.

Input: Input vector, $Z = \{z_1, z_2, ..., z_m\}$; Number of iterations, t; Primary layer, L; Number of capsules in the layer L, $C_L = \{c_1^{(L)}, c_2^{(L)}, ... , c_i^{(L)}\}$; Number of capsules in the layer L+1, $C_{L+1} = \{c_1^{(L+1)}, c_2^{(L+1)}, ... , c_j^{(L+1)}\}$

Output: CC_j , output vector

$\widehat{Z_{j|i}} = W_{ij} z_i$

$b_{ij} \leftarrow 0$, for all $c_i^{(L)} \in$ L, $c_j^{(L+1)} \in$ L+1

for *t iterations* **do**

  $\alpha_i = softmax\_AF(b_i)$ `// softmax_AF computed using (6)`

  $P_j = \sum_{i=1}^{n} \alpha_{ij} \widehat{Z_{j|i}}$, for all $c_j^{(L+1)} \in L + 1$

  $CC_j = squash\_AF(P_j)$ `// squash_AF computed using (7)`

  $b_{ij} = b_{ij} + < \widehat{Z_{j|i}}, CC_j >$

return CC_j

3.2.4 Margin loss

The output of CapsNet are class capsules, and the predictions depend on the length of the instantiation vector or norm, which represents the probability that the entity of the capsule is present or not. The class capsule with the largest instantiation vector matches the predicted class. Sabour et al. [15] proposed the classification loss or margin loss function to classify the digits on MNIST dataset. Here, a similar loss function is used to calculate the regression loss, represented as, $£_k$ and it is computed as follows:

$$£_k = IF_{k^*}(k) \max(0, m^+ - \|CC_k\|)^2 + \lambda(1 - IF_{k^*}(k)) \max(\|CC_k\| - m^-, 0)^2 \tag{8}$$

where, the values of m^+, m^-, and λ are 0.9, 0.1, and 0.5 respectively. The down−weighting factor λ, scales down the initial weight values for absent classes from affecting the model's decision. $IF_{k^*}(k)$ is an indicator function, and defined as follows:

$$IF_{k^*}(k) = \begin{cases} 0, & \text{if } k \neq k^* \\ 1, & \text{if } k = k^* \end{cases} \tag{9}$$

where, k^* is the index value of the true label. The total loss, $\check{}_T$, aggregates the loss function of all the class capsules and is computed using (10). In the training phase, an Adam optimizer is used to train the network.

$$L_T = \sum_k £_k \tag{10}$$

4 The proposed CB-Fake model

In this section, the fake news detection problem has been formulated, the steps involved in CB-Fake model and the CB-Fake algorithm for fake news detection has been discussed.

4.1 Problem formulation

The FND task can be modelled as a binary classification problem that aims to identify whether a news article in social media is fake or real news. The classification problem can be expressed as follows: Let $D = \{N_1, N_2, ..., N_m\}$ be a set of m multimedia news articles, and $y \in \{0, 1\}$ be the ground-truth label for each news $N_i \in D$. Assume that each news N_i consists of textual information (T) and visual information (I) associated with it, say $N_i = T_i \cup I_i$. Then, a model $F : D \rightarrow y$ has been defined to classify every news into the predefined class labels $y = \{0, 1\}$. It can be represented in (11) as follows:

$$F(N) = \begin{cases} 0, \text{ if N is } \textit{fake} \\ 1, \text{ if N is } \textit{real} \end{cases} \tag{11}$$

where $F(.)$ stands for the prediction function that must be learned in order to recognize fake news.

In the proposed CB-Fake model, the CapsNet and pre-trained BERT model has been introduced to efficiently capture more informative visual features and textual features. The CB-Fake model integrates these extracted features to generate high-level informative multimodal feature vector, which aids in improving the performance of the FND problem. The complete framework of the proposed CB-Fake model for fake news detection is shown in Fig. 3 and it is discussed in the upcoming subsections.

The proposed CB-Fake model is illustrated in Algorithm 2 for fake news detection. From Line 1 to Line 16, the BERT model extracts high-level textual features from the textual content of the news article. First, the textual content has been preprocessed and token ids, input mask, and segment ids has been obtained, which are represented from Line 1 to Line 6. Next, the resultant word embeddings of the sentence are then passed into an encoder part of the transformer architecture. Line 7 to Line 15 represents the sequence of steps to be followed in each encoder layer. Query (Q), Key (K) and Value (V) matrices are computed in Line 9 by performing the projection between input word vector and weight matrices W^Q, W^K and W^V. From Line 10 to Line 13, the functions of the self-attention layer and feed-forward neural layer are illustrated. Line 9 to Line 15 are repeated for the number of encoder layers, where $num\_encoder = 12$ in the BERT model. The high-level context-based textual vector, $F_{textual}$ is obtained from Line 16. The above steps are discussed in Section 3.1 in detail. CapsNet obtains informative visual features from the image content. In line 19, CapsNets performs a routing-by-agreement algorithm and returns the visual feature vector, F_{visual}. Line 20 concatenates $F_{textual}$ and F_{visual} to obtain the enhanced feature vector representation. The multimodal feature vector is fed into classification layer, which contains simple FFN with softmax activation function to predict the class label of news articles based on the probability values.

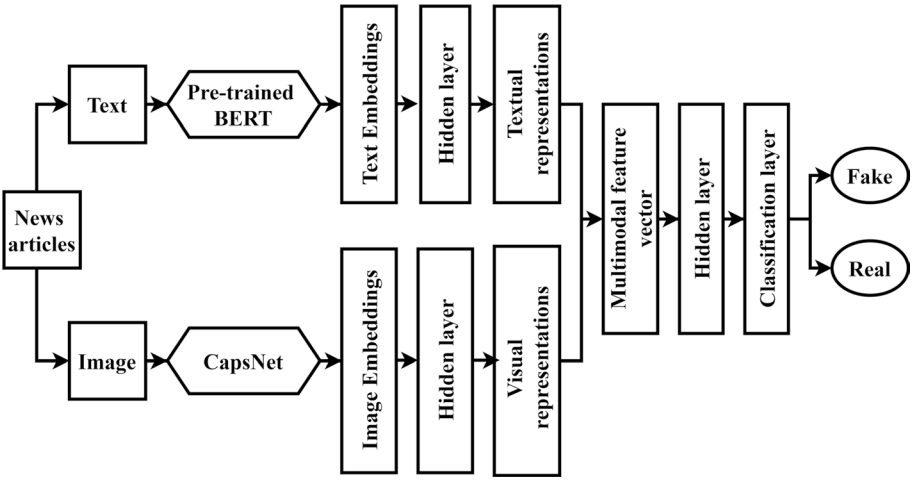


Fig. 3 Block diagram of proposed CB-Fake model for fake news detection

4.2 Data pre-processing

The FakeNewsNet repository contains two datasets, in which text and images of each news article are presented. The dataset is analysed and removed if any missing texts in the particular news id. The stop words are removed using NLTK library. The text part of the news articles are converted into a list of tokens or words, which is then transformed into a vector form for traditional machine learning models. BERT expects the input data in a specific format, which can be explained in the preceding paragraphs. The feature vector of an image of the news can be represented by CapsNet model.

4.3 Feature extraction

The key idea of feature extraction is to extract the meaningful information from the multimodal news articles. In this work, BERT-base-uncased and CapsNet model are used to extract the important features from the textual and the visual content of the news, which is discussed in the Sections 3.1 and 3.2 respectively.

4.3.1 Textual feature vector representation

The computation steps of BERT for an input sequence are discussed as follows: At first, BERT tokenizes given sentences into word pieces, and then a fixed-length vector of dimension 128 is obtained by combining the vectors of three embedding layers namely token, position, and segment. The special token [CLS] and [SEP] has been

added to distinguish the start and end of the sentence, as depicted in Fig. 4. The final word embedding vector is passed to the encoder module of the transformer. BERT-base model has 12 encoder layers. For each word vector, attention matrix is calculated using (1) and the (2) concatenates all these attention matrices. After the eight multihead attention blocks, the resultant vector is sent to the FFN in parallel. The output vector of this FFN layer is passed from the present encoder to the next encoder. This procedure is repeated twelve times since the total number of an encoder is 12 in this BERT model. Finally, the context-based word embedding vector of dimension 768 is obtained from the transformer encoder module. This vector is fed into a dropout layer with 0.2 probability ratio and then passed to two fully connected layers of dimensions 768 and 32. In BERT training phase, the cross-entropy loss function is used as objective function and an Adam optimizer is utilized to train the model. These steps are used in Algorithm 2 to extract textual feature vector of dimension 32, namely $F_{textual}$. The semantic-based textual feature representation $F_{textual}$ will be concatenated with the visual feature vector F_{visual} to obtain an enhanced multimodal feature vector, F_{tv} for fake news classification.

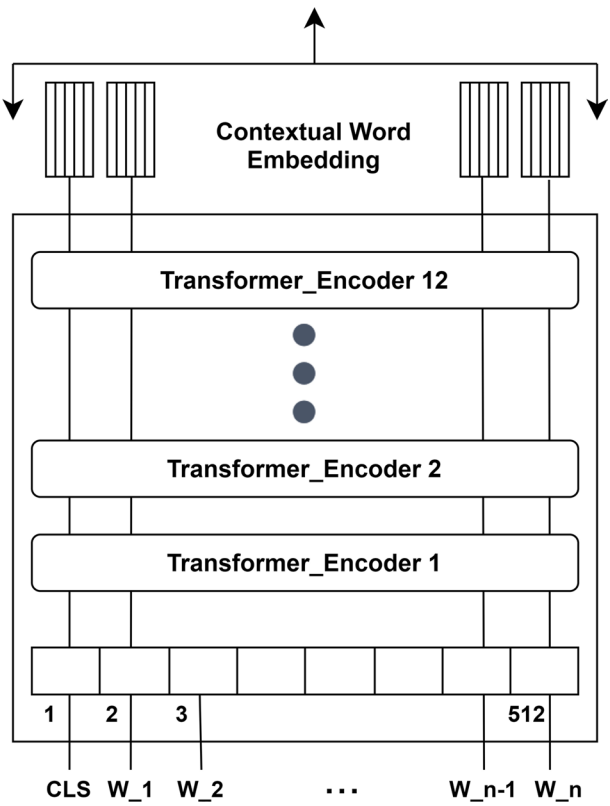


Fig. 4 A high-level diagram of textual feature representation using BERT

4.3.2 Visual feature extraction and representation

The visual features are useful while the textual information is wrong about the original fake news in predicting fake or real news. Hence, it is necessary to extract important visual features from an image of the multimodal news article. In this work, CapsNet is used for visual feature extraction. The steps involved in each CapsNet layer to obtain the visual feature vector of dimension 32, F_{visual} are explained in Section 3.2.

4.4 Concatenation of feature vectors

The multimodal feature vector F_{tv} is obtained by concatenating textual $F_{textual}$ and visual feature vector F_{visual} of dimension 32 using the 50-50 weightage method for text and image features. Concatenation means the average of the vector values is computed for each position vector. The model also uses 60-40 and 40-60 weightage methods for multimodal feature vector representation. In experiments, it is observed that the equal weightage method produces promising results than the other aforementioned weightage methods. The resultant vector of dimension 32, is passed to the classification layer.

4.5 Classification

The concatenated multimodal feature vector of dimension 32 is fed into this layer. The simple feed forward neural network with softmax activation function is employed to predict the fake news based on the predicted probability values. In this problem, the labels assigned to fake and real are 0 and 1 respectively. If the probability value of a news article is closer to 0, then the model predicted as fake news articles, otherwise (closer to 1), predicted as true news.

Algorithm 2 The proposed CB-Fake algorithm.

Input: Set of news articles, $D = \{N_1, N_2, ..., N_m\}$, where $N_q = T_q \cup I_q$, $q = 1, 2, ...m$; Set of textual contents, $T = \{T_1, T_2, ..., T_m\}$; Set of visual contents, $I = \{I_1, I_2, ..., I_m\}$; Number of encoder layer, $num\_encoder = 12$; Number of heads, $num\_heads = 8$

Output: *class_label*, $\{fake(0), real(1)\}$

```
/* Textual feature vector representation using BERT model
   */
```

Compute the following steps to convert the textual input data $T = \{T_1, T_2, ..., T_m\}$ into word embeddings

for *textual_content in T* **do**

  for *sentence in textual_content* **do**

    Tokenize sentence and map the tokens to vocabulary IDs

    Set the maximum sequence length sentence

    Create attention mask IDs and Segment IDs

Pass the final input word embeddings into the Encoder part of Transformer model

for *1 ... num_encoder* **do**

  Obtain matrices Q, K, and V for each preprocessed word embedding with corresponding initial weight matrices W^Q, W^K, W^V

  for *1 ... num_heads* **do**

    for *word in sentence* **do**

      Calculate the scaled dot product attention for each word in a sentence using (2)

      Compute multi-head attention using (3)

  Resultant vector from the step 10 is then passed to the feed forward neural network (FFN) layer concurrently

  Obtain output vector from FFN layer of the current encoder and pass it to the next encoder layer

Obtain final context-based textual features, $F_{textual}$ from the step 8

```
/* Visual feature vector representation using CapsNet
   architecture                                        */
```

for *visual content in I* **do**

  Obtain the vector representation $Z = \{z_1, z_2, ..., z_m\}$

Obtain visual feature vector, F_{visual} using routing-by-agreement algorithm `// Call Algorithm 1`

```
/* Textual and Visual feature vector concatenation      */
```

Obtain rich feature vector representation, $F_{tv} = F_{textual} \oplus F_{visual}$

Predict the class label for F_{tv} using simple FFN with softmax activation function

return *class_label*

5 Experimental setup

In this work, the performance of the proposed CB-Fake model is systematically evaluated on Politifact and Gossipcop datasets based on the accuracy measure. The effectiveness of the model is compared with the classical ML models, ensemble techniques and

state-of-the-art methods. Finally, the limitations of the proposed model are observed based on the above experiments.

5.1 Datasets

A publicly available comprehensive dataset termed FakeNewsNet provided by the author Kai Shu [40] is used for our work. It contains two datasets, namely Politifact and Gossipcop, which consists of news articles related to politics and entertainment respectively. The proposed CB-Fake model is analysed using these two datasets to evaluate the effectiveness of this model. The dataset comprises of news articles, with each article having text and an image associated with it. The ground-truth labels for the political and entertainment domain were collected from Politifact[4], Gossipcop[5] and E! Online[6], respectively. The description of the preprocessed dataset is given in Table 7. After data preprocessing, the suitable samples are selected and used in our experiments, which is indicated in the square brackets.

The dataset is divided into training and testing in the 70:30 ratio, respectively and is shown in Table 8.

5.2 Baseline models

The proposed CB-Fake model is experimented with benchmark datasets and compared with three base modalities that focus on machine learning, single modality, and multimodal model for fake news identification.

5.2.1 Machine learning models

- *NB* [31]: Multinomial NB is a classic machine learning model used for classification of text documents with the help of count of words as vectors.
- *SVM* [1, 33]: SVM identify fake information from news documents by forming the linguistic attributes. Typically a linear kernel is used.
- *RF* [9]: RF is a meta estimator that is used to fit multiple decision tree classifiers on the dataset which improves the predictive accuracy and reduces the overfitting.
- *SGD*: SGD is used to model unconstrained problems that leads to an optimized result. It calculates the gradient of the error one training sample at a time and updates the parameters of the learning function.
- *LR* [43]: LR is used to formulate the article content. The document is vectorized using the TF-IDF method and then it is classified.
- *Decision fusion approach*: The voting classifier combines two or more base classifiers and applies majority voting to predict the class labels.

[4] https://www.politifact.com/
[5] https://www.gossipcop.com/
[6] https://www.eonline.com/ap

Table 7 The statistics of the FakeNewsNet dataset

Dataset	Politifact	Gossipcop
Real News	624 [499]	16817 [15223]
Fake News	432 [376]	5323 [4784]

Table 8 The details of training and testing data

Details	Politifact	Gossipcop
Total samples (TS)	875	20,007
Training data (70% x TS)	612	14,004
Testing data (30% x TS)	263	6,003

5.2.2 Single modality models

- Textual
  - *CNN* [55]: Convolutional neural networks are layers of convolutions with nonlinear activation functions applied to the results. The simple FFN with softmax activation function has been used for classifying a news article.
  - *XLNet* [43]: It is a generalized auto regressive pre-training method which used the context word to predict the next word. The context is constrained in both forward and backward directions.
  - *LSTM* [47]: It is a recurrent neural network model which is used in learning scenarios where dependencies between inputs has to be preserved.
- Visual
  - *VGG19* [20, 43, 44, 49, 53]: Images are fed as input to this model which extracts visual features, which is then passed to a fully connected layer to identify fake news articles.

5.2.3 Multimodal models

- *EANN* [49]: EANN is a prominent multimodal framework for fake information detection on news articles. The feature extractor module of EANN utilizes Text-CNN and VGG-19 network to obtain visual and textual features from social media. These features are later used to train the EANN for generating the classifier model to understand the respective news is either fake or real. The last component, event discriminator is accountable for eliminating the event-specific features. In our experiments, we use EANN, which excludes the event discriminator part for a fair comparison.
- *MVAE* [20]: MVAE framework computes the correlations between various modalities using variational auto encoder (VAE). This helps to reconstruct the visual and textual features from a shared representation.
- *SpotFake* [44]: The SpotFake framework, a transfer learning based fake information detection model, aims to learn textual and visual features by BERT and VGG-19 model, respectively.

- *SpotFake+* [43]: SpotFake+ employs XL-Net model for textual feature extraction and VGG-19 network for extracting visual features.

5.3 Parameter setup

The experiments are carried out on the Google Colab platform, used to build, train, test, and assess the models. The configuration specifications are as follows: CPU: 1x Intel(R) Xeon(R) CPU @2.3GHz, GPU: 1x Tesla K80, 2496 CUDA cores, 12GB GDDR5 VRAM, RAM size: 12.6GB, and Disk size: 33GB. Python version 3.6.4 is used to implement all the codes. The dataset is first preprocessed. Split the dataset into training and testing in a 70:30 ratio using the "Hold and out" cross-validation approach. The proposed CB-Fake model and traditional machine learning models are implemented using the TensorFlow (1.15) and scikit-learn libraries. For text preprocessing, the NLTK library and CountVectorizer are used. The BERT Model is trained for 10 epochs on 70% training data (textual content) and validated on test data. The cross-entropy loss function and AdamW optimizer are used to train the BERT model, with a learning rate of $2e^{-5}$ and an epsilon of $1e^{-8}$, a maximum sequence length of 128 tokens, and a batch size of 32. BERT model uses the hidden layer of dimension 768. Therefore, the output vector of dimension 768 is obtained after the word embeddings process. It is then passed to the fully connected layer and drop out layer of size 768 and 32 respectively. Finally, the context-based textual feature vector $F_{textual}$ of dimension 32 is obtained from BERT.

Furthermore, CapsNet is used to generate the visual feature vector from the image of the news article. It consists of three layers, with hyperparameters for each layer mentioned in Table 9. In the training process, CapsNet uses batch_size = 100 and num_epochs = 30. In the primary capsule layer (PCL), we have used eight child capsules and two-parent capsules in CCL. The complexity of the routing-by-agreement algorithm depends on the number of capsules and the number of intermediate higher capsule layers was used. Depends on these, the total number of hyperparameters will vary, which is smaller than CNN. The capsule connections are made between a group of neurons rather than individual neurons in the CapsNet model, thus it requires fewer parameters than CNN. Thus, the CapsNet model takes minimum time to train all the data samples compared to CNN. We have used the trial and error method, and then fixed only one higher-capsule layer which is used to obtain the visual feature vector, F_{visual} of dimension 32. The proposed model then combines the textual and visual feature vectors of dimension 32 to generate a high-level informative multimodal feature vector, F_{tv} of dimension 32. Finally, the fully connected softmax layer receives this fused vector, which is used to classify fake and real news based on the predicted probability values.

Table 9 Hyperparameters of CapsNet layers for visual feature representation

Layer	Num_Capsules	Num_routes	In_channels	Out_channels	Kernel_size
COL	–	–	1	256	9
PCL	8	–	256	32	9
CCL	2	32 * 6 * 6	8	16	–

5.4 Evaluation metrics

We used the traditional performance metrics namely accuracy, recall, precision and F1-score has been calculated using the equations from (12) to (15) to evaluate the proposed framework. Furthermore, the fake news identification problem is treated similar to a classification problem that classifies whether a news article is fake or real. The confusion matrix is used to compute the performance of the fake news detection. A brief explanation of these metrics is as follows:

$$Accuracy = \frac{TP + TN}{TP + TF + FP + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{15}$$

where, True positive (TP) = Fake news predicted as fake; True negative (TN) = Real news predicted as real; False positive (FP) = Fake news predicted as real; False negative(FN) = Real news predicted as fake.

In the fake news identification problem, the accuracy value shows the fraction of correctly classified news to the number of news articles. Precision is the ratio of correctly classified fake news with the total number of predicted fake news articles. Recall or true positive rate (TPR) metric is calculated as the ratio of fake news articles correctly classified as fake to the total number of actual fake news articles. The F1-score metric is the harmonic mean value of the precision and recall obtained for the fake news identification and thus provides an overall performance of the proposed approach.

6 Results and discussions

The experimental results of the proposed model has been analysed in detail with different performance evaluation metrics and elaborative discussion has been done in the preceding paragraphs.

The proposed CB-Fake model is compared with the benchmark dataset named FakeNewsNet. The obtained results are shown in Tables 10, 11 and 12. From the experiment results, it is observed that the proposed CB-Fake shows better performance than the other traditional methods, in terms of accuracy, precision, recall and F1-score.

The high accuracy and better F1-score signifies the fake news detection capability of BERT model on textual features is shown in Table 10. For Politifact dataset, the BERT model achieves 29-32% and 6-7% improvements over traditional classifiers (NB, SVM) and ensemble classifier in terms of accuracy. Similarly, in F1-score metric, BERT model obtains 11-29% improvements compared to all other classifiers and is shown in Fig. 5a. For Gossipcop dataset, BERT achieves 29-42% and 6-10% accuracy gain over base classifiers

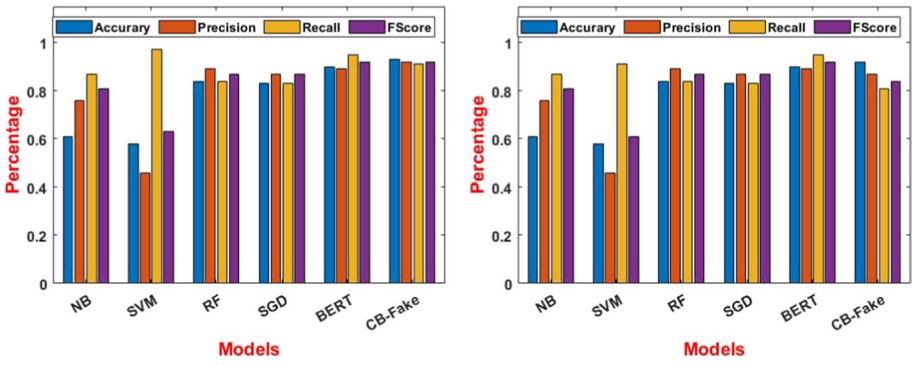
Table.10 Comparison of BERT model with base classifier on textual features of dataset

Classifier	Politifact				Gossipcop			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
NB [31]	0.61	0.76	0.87	0.81	0.62	0.79	0.91	0.85
SVM [1, 33]	0.58	0.46	0.91	0.61	0.49	0.46	0.91	0.61
RF [9]	0.84	0.89	0.84	0.87	0.85	0.98	0.85	0.91
SGD	0.83	0.87	0.83	0.85	0.81	0.88	0.87	0.87
BERT	**0.90**	0.89	0.95	**0.92**	**0.91**	0.92	0.97	**0.94**
CB-Fake	**0.93**	0.92	0.91	**0.92**	**0.92**	0.87	0.81	0.84

(Maximum accuracy and F1-Score are shown in bold)

and ensemble classifier Fig. 5b. BERT model outperforms the other traditional classifier and ensemble classifier, as shown in Fig. 5. The following are the main reasons for the BERT model's superiority over other conventional classifiers: it generates high-level contextual embedding between words in a sentence of a news article using the self-attention mechanism, parallelization in the training phase, and bidirectional accessing of words in a sentence. Since our proposed CB-Fake model is the fusion of BERT and CapsNet, it incorporates the context-based textual features obtained from BERT along with the informative visual features from CapsNet. Hence it has achieved higher accuracy of 93% and 92%, which is 32-35% and 12-13% higher than the traditional ML classifiers and 3% & 1% higher than BERT for Politifact and Gossipcop dataset respectively. Similarly, the proposed CB-Fake model has better performance in terms of precision, recall and F1-score for Politifact dataset.

Further, the proposed model is compared with decision fusion approach, especially voting classifier and the results are shown in Table 11. BERT model achieves better accuracy, recall and F1-score than the voting classifier, as shown in Fig. 6. For Politifact dataset, the combination of NB, RF, SGD achieves 87% and 89% for accuracy and F1-score respectively which is higher than the other voting classifier combinations,



(a) Politifact dataset (b) Gossipcop dataset

Fig. 5 The performance of the BERT model and base classifiers

which is shown in Fig. 6a. Figure 6b depicts that the voting classifier with NB, SVM and RF performs better than other ensemble models for Gossipcop dataset. Politifact dataset size is small when compared to the Gossipcop dataset. Due to this size of the dataset, the performance of the different combinations of the voting classifier is varying. But when compared to the BERT model, the voting classifier's prediction accuracy is less for both datasets. BERT achieves 7% accuracy gain on an average for both Poltifact and Gossipcop datasets. The improvement in accuracy inferred that the contextual word embeddings of BERT identifies the fake news effectively. But our CB-Fake model has achieved better accuracy than BERT for both Politifact and Gossipcop datasets. From this, we can conclude that CB-Fake outperforms other ensemble classifiers which is evident from Table 11.

The confusion matrix of the proposed CB-Fake model on test data (30% data) on Politifact and Gossipcop dataset is shown in Fig. 7.

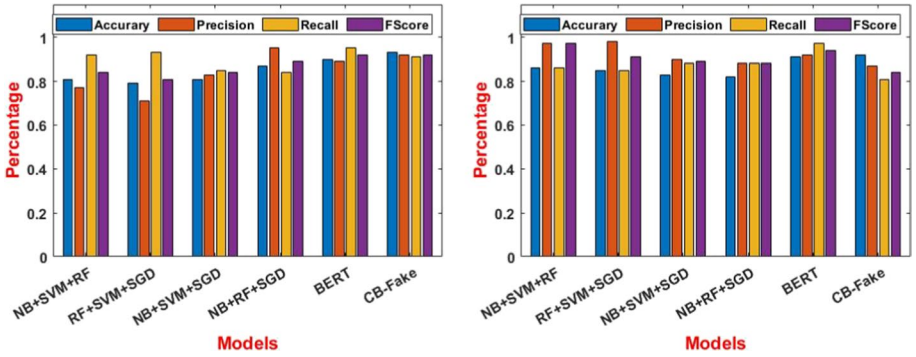
Table 12 includes the results of classification on Politifact and Gossipcop datasets with our model and traditional classifiers, is depicted Fig. 8. We selected the accuracy measure which is used in standard methods as the metric for performance evaluation. For Politifact dataset, the BERT model works better than the other single modality textual models. It achieves 8.7% accuracy gain on an average over three XLNet models and three base classifiers. In addition, 27.1% accuracy gain was achieved with BERT over the CNN model. Hence, the single modality based BERT model shows better performance on fake news identification. Among other multimodal models, our CB-Fake model improves the accuracy than all the other baseline methods, as shown in Fig. 8a. This significant improvement of our CB-Fake model depends on the self-attention function of the BERT model and routing-by-agreement procedure of CapsNet architecture. The accuracy gain also signifies the capability of BERT model. Though textual features are more suited for fake news experiments, there are some concerns on textual features compared to the visual features in unimodality mode. Specifically, when compared to the recent baseline model SpotFake+ and SpotFake, our proposed CB-Fake outperforms both models by 8.4% and 20.9%, respectively.

Furthermore, in the Gossipcop dataset, similar nature of output and performance has been found. A significant accuracy difference is observed for textual features compared to the visual part, as shown in Fig. 8b. The average performance gain over XLNet models and classical machine learning models is 8.23% and 32.03%. Moreover, CB-Fake achieves an 18.7% accuracy gain over the CNN model because our proposed model eliminates the

Table 11 Comparison of BERT model with decision fusion classifier on textual features of dataset

Classifier	Politifact				Gossipcop			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
NB+SVM+RF	0.81	0.77	0.92	0.84	0.86	0.97	0.86	0.91
RF+SVM+SGD	0.79	0.71	0.93	0.81	0.85	0.98	0.85	0.91
NB+SVM+SGD	0.81	0.83	0.85	0.84	0.83	0.90	0.88	0.89
NB+RF+SGD	0.87	0.95	0.84	0.89	0.82	0.88	0.88	0.88
BERT	**0.90**	0.89	0.95	**0.92**	**0.91**	0.92	0.97	**0.94**
CB-Fake	**0.93**	0.92	0.91	**0.92**	**0.92**	0.87	0.81	0.84

(Maximum accuracy and F1-Score are shown in bold)



(a) Politifact dataset (b) Gossipcop dataset

Fig. 6 The performance of the BERT model and decision fusion classifier

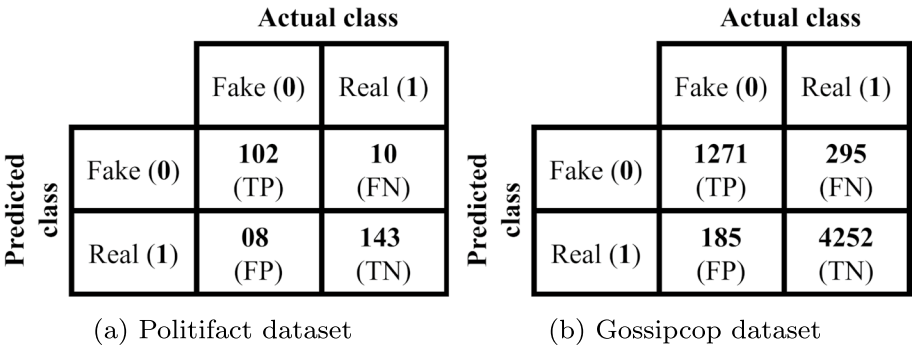
loss of information due to pooling operation. Among multimodal models, the proposed CB-Fake performs better than all the other baselines. CB-Fake achieves 11.3% and 6.4% accuracy gain over SpotFake and SpotFake+, respectively. The reason for our proposed CB-Fake model works better than the other state-of-the-art methods is because of the context-based textual vector captured by BERT's multihead attention process and the highly informative visual features extracted by using CapsNet's routing-by-agreement algorithm. From these results, it is witnessed that the proposed CB-Fake model is effective in identifying the fake news from various social media platforms.

Although the CB-Fake model outperformed all other approaches, it has few limitations which are discussed as follows. The proposed model takes textual and visual features from news articles as input; however the user profile information and user's behavioral characteristics were not analysed. Also, as the CB-Fake model is trained using fake news datasets of english language, it cannot be used for detecting other language datasets.

Table 12 The performance of the proposed CB-Fake model with baselines on FakeNewsNet dataset

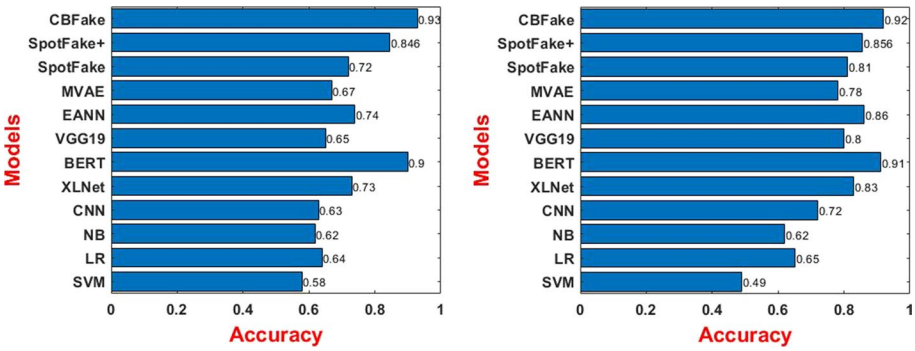
Modality	Models	Politifact	Gossipcop
Textual	SVM [1, 33]	0.58	0.497
	LR [43]	0.642	0.648
	NB [31]	0.617	0.624
	CNN [55]	0.629	0.723
	XLNet + dense layer [48]	0.74	0.836
	XLNet + CNN [48]	0.721	0.84
	XLNet + LSTM [48]	0.721	0.807
	BERT	**0.90**	**0.91**
Visual	VGG19 [20, 43, 44, 49, 53]	0.654	0.80
Multimodal (Textual+Visual)	EANN [49]	0.74	0.86
	MVAE [20]	0.673	0.775
	SpotFake [44]	0.721	0.807
	SpotFake+ [43]	0.846	0.856
	CB-Fake	**0.93**	**0.92**

(Maximum accuracy is shown in bold)



(a) Politifact dataset (b) Gossipcop dataset

Fig. 7 Confusion matrix results of the proposed CB-Fake model on testing data



(a) Politifact dataset (b) Gossipcop dataset

Fig. 8 The performance of the proposed CB-Fake model with state-of-the-art methods

7 Conclusion

In this work, we developed an end-to-end model for fake news identification at the early stage by analyzing both the textual and visual content of the news article. The significant limitations of the current models for the FND task are the lack of extracting informative features from the text and its associated image of the multimedia news. To address this issue, the proposed CB-Fake model aggregates textual and visual features to learn an enhanced multimodal feature representation. The CB-Fake model incorporates CapsNet for most informational visual feature extraction from the image. It also employs a pre-trained language model BERT to capture strong context-based textual features from the news articles. These features are then combined to create a richer data representation, which is then sent to a classification layer to determine whether the news is fake or real. The proposed model's performance is evaluated using two publicly available datasets obtained from social networking sites. Among the multimodal fake news detection model, the experimental results show that the proposed model, CB-Fake, is efficient and outperforms current state-of-the-art methods. Especially, it achieves better accuracy than the recent baseline model, SpotFake+, by a margin of 8% on an average for both fake news datasets.

In the future, different deep learning models for fusing textual and visual features can be investigated to better understand the relationship between different modalities to recognize fake news. In addition, the proposed model would use social-context features such as user profile information, propagation patterns, etc., for effective fake news prediction. The proposed CB-Fake model only analyzes English fake news datasets for identifying if the news is fake or not, but it can be extended to include other popular languages' fake news datasets. Further, the proposed model can be also used to detect the fake news related to COVID-19, thereby enhancing the public knowledge about these pandemic diseases with proper information.

References

1. Ahmed H, Traore I, Saad S (2017) Detection of online fake news using n-gram analysis and machine learning techniques. International conference on intelligent, secure, and dependable systems in distributed and cloud environments. Springer, Cham, pp 127–138
2. Akyol K, Sen B (2019) Modeling and predicting of news popularity in social media sources. Cmc-Computers Materials & Continua 61(1):69–80
3. Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. Journal of economic perspectives 31(2):211–36
4. Antol S, Agrawal A, Lu J, Mitchell M, Batra D, Zitnick CL, Parikh D (2015) Vqa: Visual question answering. In: Proceedings of the IEEE international conference on computer vision, pp 2425–2433
5. Asghar MZ, Habib A, Habib A, Khan A, Ali R, Khattak A (2019) Exploring deep neural networks for rumor detection. J Ambient Intell Human Comput 12:4315–4333
6. Bondielli A, Marcelloni F (2019) A survey on fake news and rumour detection techniques. Information Sciences 497:38–55
7. Chen T, Li X, Yin H, Zhang J (2018) Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. Pacific-Asia conference on knowledge discovery and data mining. Springer, Cham, pp 40–52
8. Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805
9. Faustini PHA, Covões TF (2020) Fake news detection in multiple platforms and languages. Expert Systems with Applications 158:113503
10. Goldani MH, Momtazi S, Safabakhsh R (2021) Detecting fake news with capsule neural networks. Applied Soft Computing 101:106991
11. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al (2014) Generative adversarial nets. Adv Neural Inf Process Syst 27:2672–2680
12. Guo B, Ding Y, Yao L, Liang Y, Yu Z (2020) The Future of False Information Detection on Social Media: New Perspectives and Trends. ACM Computing Surveys (CSUR) 53(4):1–36
13. Guo C, Cao J, Zhang X, Shu K, Liu H (2019) Dean: Learning dual emotion for fake news detection on social media (arXiv preprint). arXiv:1903.01728
14. Gupta M, Zhao P, Han J (2012) Evaluating event credibility on twitter. In: Proceedings of the 2012 SIAM international conference on data mining, society for industrial and applied mathematics, pp 153–164
15. Hinton GE, Sabour S, Frosst N (2018). Matrix capsules with EM routing. In: International conference on learning representations
16. Hinton GE, Krizhevsky A, Wang SD (2011) Transforming auto-encoders. International conference on artificial neural networks. Springer, Berlin, pp 44–51
17. Jin Z, Cao J, Guo H, Zhang Y, Luo J (2017) Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In: Proceedings of the 25th ACM international conference on multimedia, pp 795–816
18. Kaliyar RK, Goswami A, Narang P (2021) FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia Tools and Applications 80(8):11765–11788
19. Karpathy A, Fei-Fei L (2015) Deep visual-semantic alignments for generating image descriptions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3128–3137
20. Khattar D, Goud JS, Gupta M, Varma V (2019) Mvae: Multimodal variational autoencoder for fake news detection. In: The world wide web conference, pp 2915–2921

21. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, ... Baddour K (2020) Corona virus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter. Cureus 12(3)
22. Kumar A, Sangwan SR, Nayyar A (2019) Rumour veracity detection on twitter using particle swarm optimized shallow classifiers. Multimedia Tools and Applications 78(17):24083–24101
23. Liu Y, Wu YF (2018) Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Proceedings of the AAAI conference on artificial intelligence, vol 32, no 1
24. Li Q, Zhang Q, Si L (2019) Rumor detection by exploiting user credibility information, attention and multi-task learning. In: Proceedings of the 57th annual meeting of the association for computational linguistics, pp 1173–1179
25. Ma J, Gao W, Mitra P, Kwon S, Jansen BJ, Wong KF, Cha M (2016) Detecting rumors from microblogs with recurrent neural networks
26. Ma J, Gao W, Wong KF (2018) Detect rumor and stance jointly by neural multi-task learning. In: Companion proceedings of the the web conference, pp 585–593
27. Ma J, Gao W, Wong KF (2018) Rumor detection on twitter with tree-structured recursive neural networks. Association for Computational Linguistics
28. Marra F, Gragnaniello D, Cozzolino D, Verdoliva L (2018) Detection of gan-generated fake images over social networks. In: 2018 IEEE Conference on multimedia information processing and retrieval (MIPR), IEEE, pp 384–389
29. Nayak P (2019) Understanding searches better than ever before, available at: https://www.blog.google/products/search/search-language-understanding-bert/
30. Ozbay FA, Alatas B (2019) A novel approach for detection of fake news on social media using metaheuristic optimization algorithms. Elektronika ir Elektrotechnika 25(4):62–67
31. Ozbay FA, Alatas B (2020) Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A: Statistical Mechanics and its Applications 540:123174
32. Patrick MK, Adekoya AF, Mighty AA, Edward BY (2019) Capsule network–a survey. J King Saud Univ-Comput Inf Sci
33. Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R (2017) Automatic detection of fake news. arXiv:1708.07104
34. Qi P, Cao J, Yang T, Guo J, Li J (2019) Exploiting multi-domain visual information for fake news detection. In: 2019 IEEE International conference on data mining (ICDM), IEEE, pp 518–527
35. Rapoza K (2017) Can fake news impact the stock market? https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/. 26 February
36. Roger M (2019) Google's BERT rolls out worldwide, available at: https://www.searchenginejournal.com/google-bert-rolls-out-worldwide/339359/
37. Sabour S, Frosst N, Hinton GE (2017) Dynamic routing between capsules. arXiv:1710.09829
38. Savyan PV, Bhanu SMS (2020) UbCadet: detection of compromised accounts in twitter based on user behavioural profiling. Multimedia Tools and Applications 79:1–37
39. Shu K, Sliva A, Wang S, Tang J, Liu H (2017) Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter 19(1):22–36
40. Shu K, Mahudeswaran D, Wang S, Lee D, Liu H (2020) Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data 8(3):171–188
41. Shu K, Cui L, Wang S, Lee D, Liu H (2019) Defend: Explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp 95–405
42. Singh S, Cha J, Kim TW, Park J (2021) Machine learning based distributed big data analysis framework for next generation web in IoT. Comput. Sci. Inf. Syst. 18:597–618
43. Singhal S, Kabra A, Sharma M, Shah RR, Chakraborty T, Kumaraguru P (2020) Spotfake+: A multimodal framework for fake news detection via transfer learning (student abstract). In Proceedings of the AAAI Conference on Artificial Intelligence 34(10):13915–13916
44. Singhal S, Shah RR, Chakraborty T, Kumaraguru P, Satoh SI (2019) Spotfake: A multi-modal framework for fake news detection. In: 2019 IEEE Fifth International conference on multimedia big data (BigMM), IEEE, pp 39–7
45. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, ... & Polosukhin I (2017) Attention is all you need. arXiv:1706.03762
46. Vesperini F, Gabrielli L, Principi E, Squartini S (2019) Polyphonic sound event detection by using capsule neural networks. IEEE Journal of Selected Topics in Signal Processing 13(2):310–322
47. Wang WY (2017) Liar, liar pants on fire: A new benchmark dataset for fake news detection. arXiv:1705.00648

48. Wang Y, Huang L, Jiang S, Wang Y, Zou J, Fu H, Yang S (2020) Capsule networks showed excellent performance in the classification of hERG blockers/nonblockers. Frontiers in pharmacology 10:1631
49. Wang Y, Ma F, Jin Z, Yuan Y, Xun G, Jha K, ... & Gao J (2018) Eann: Event adversarial neural networks for multi-modal fake news detection. In: Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining, pp 849–857
50. Wu L, Liu H (2018) Tracing fake-news footprints: Characterizing social media messages by how they propagate. In: Proceedings of the eleventh ACM international conference on web search and data mining, pp 637–645
51. Wu K, Yang S, Zhu KQ (2015) False rumors detection on sina weibo by propagation structures. In: 2015 IEEE 31st International conference on data engineering, IEEE, pp 651–662
52. Yang K, Long S, Zhang W, Yao J, Liu J (2020) Personalized News Recommendation Based on the Text and Image Integration. CMC-Computers Materials & Continua 64(1):557–570
53. Yang Y, Zheng L, Zhang J, Cui Q, Li Z, Yu PS (2018) TI-CNN: Convolutional neural networks for fake news detection. arXiv:1806.00749
54. Yin L, Meng X, Li J, Sun J (2019) Relation extraction for massive news texts. Comput Mater Continua 58:275–285
55. Yu F, Liu Q, Wu S, Wang L, Tan T (2017) A convolutional approach for misinformation identification, In: IJCAI, pp 3901–3907
56. Zeng J, Ma X, Zhou K (2019) Photo-realistic face age progression/regression using a single generative adversarial network. Neurocomputing 366:295–304
57. Zhou X, Zafarani R (2020) A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR) 53(5):1–40
58. Zhou X, Jain A, Phoha VV, Zafarani R (2020) Fake news early detection: A theory-driven model. Digital Threats: Research and Practice 1(2):1–25
59. Zhou P, Han X, Morariu VI, Davis LS (2018) Learning rich features for image manipulation detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1053–1061
60. Zhou X, Zafarani R, Shu K, Liu H (2019) Fake news: Fundamental theories, detection strategies and challenges. In: Proceedings of the twelfth ACM international conference on web search and data mining, pp 836-837

Authors and Affiliations

Balasubramanian Palani[1] · Sivasankar Elango[1] · Vignesh Viswanathan K[2]

Sivasankar Elango
sivasankar@nitt.edu

Vignesh Viswanathan K
vigneshkvn2098@gmail.com

[1] Department of Computer Science and Engineering, National Institute of Technology, Tiruchirappalli, India

[2] Software Development Engineer Visa Inc, Bengaluru, India