FakedBits- Detecting Fake Information on Social Platforms using Multi-Modal Features

Dilip Kumar Sharma[1], Bhuvanesh Singh[2], Saurabh Agarwal[3]*, Hyunsung Kim[4]* and Raj Sharma[5]

[1] GLA University, Mathura, India
[2] University of St. Thomas, Minnesota, U.S.A.
[3] Amity School of Engineering & Technology, Amity University Uttar Pradesh, Noida, India
[4] School of Computer Science, Kyungil University, Gyeongsan, South Korea
[5] SRM Institute of Science and Technology, Chennai, India
[e-mail : dilip.sharma@gla.ac.in, bhuvaneshsingh80@gmail.com, saurabhnsit2510@gmail.com, kim@kiu.ac.kr, rs9602@srmist.edu.in]
*Corresponding Authors: Saurabh Agarwal, Hyunsung Kim

Received July 25, 2022; revised September 22, 2022; revised November 22, 2022; accepted January 17, 2023; published January 31, 2023

Abstract

Social media play a significant role in communicating information across the globe, connecting with loved ones, getting the news, communicating ideas, etc. However, a group of people uses social media to spread fake information, which has a bad impact on society. Therefore, minimizing fake news and its detection are the two primary challenges that need to be addressed. This paper presents a multi-modal deep learning technique to address the above challenges. The proposed modal can use and process visual and textual features. Therefore, it has the ability to detect fake information from visual and textual data. We used EfficientNet-B0 and a sentence transformer, respectively, for detecting counterfeit images and for textural learning. Feature embedding is performed at individual channels, whilst fusion is done at the last classification layer. The late fusion is applied intentionally to mitigate the noisy data that are generated by multi-modalities. Extensive experiments are conducted, and performance is evaluated against state-of-the-art methods. Three real-world benchmark datasets, such as MediaEval (Twitter), Weibo, and Fakeddit, are used for experimentation. Result reveals that the proposed modal outperformed the state-of-the-art methods and achieved an accuracy of 86.48%, 82.50%, and 88.80%, respectively, for MediaEval (Twitter), Weibo, and Fakeddit datasets.

Keywords: Deep learning; Learning systems, Social networking sites, Fake information, Multi-modal

1. Introduction

There is a cultural change in how information is processed nowadays by individuals. To collect more information quickly, people often search for summarized copies on social media sites [1]. Taking advantage of this reliance, social networking sites are used to distribute fake news. Fake news has been circulating for a long time. Neogi et al. [2] observed that it is now frequently used to spread a narrative or propaganda under the guise of politics. Norwegian media authority studied false information about the coronavirus in March 2020 [3]. The study revealed that social platforms, specifically social networking websites, held a significant contribution to the dissemination of incorrect information. Bunker [4] named this fake information (about Covid-19) as an infodemic. He claimed that it was more dangerous than the Covid-19 virus. Fake pictures and videos were used extensively in the dissemination of fake news. Fazio [5] found that false pictures attract three times greater attention than text. Fake images are digitally modified images that are subject to several changes. These days, fake news has become a societal threat. Some consequences of false photos and recordings have been serious, such as mob lynching. Thus, it is essential to detect and curb fake news on social networking sites.

The fake news spread with malicious intent has a significant impact on society. Counterfeit images are used widely to provoke anger and polarize people's emotions. However, it has a more negative impact on severe consequences such as religious feuds, mob lynching's improper patient care recommendations. According to a poll performed by CIGI-IPSOS and the Internet Society [6], the top two sites for disseminating false news are Facebook and Twitter. A similar observation was also reported in [7]. According to a poll, Facebook is the source of the majority of public health concerns through disseminating substantial health misinformation [8]. As a result, developing methods to detect fake news on these social networking sites is essential. Sharma and Sharma [9] present several ways of detecting it. A similar review of various techniques to detect rumors over social network platforms was presented in [10]. To detect false news, a multi-modal method was recently utilized. Textual, images, network propagation, statistical cues, and user profiles are some of the content and context kinds used in multi-modal frameworks.

For fake news identification, we propose a multimodal method that uses the latest and updated architecture to detect fake images and false text. Multi-modal architecture opts because it works better than any singular modular framework [11]. Also, as suggested in [5], images play a major role than text in spreading fake news. This paper presents the usage of distilled bidirectional encoder-based Sentence Transformer (distilBERT) [12] for text feature learning and deep convolutional neural network (CNN) model EfficientNet for image feature learning [13]. Bidirectional encoder representations from transformers" (BERT) is used widely in text classifications. distilBERT has shown superior results than the BERT-base with small parameters for the online news dataset MRPC [12]. On the other hand, CNN's have been used for image processing, identification, and classifications. There are several well-known CNN models. such as ResNet, InceptionNet, and ImageNet are available for image classification. Literature [13] revealed that the EfficientNetB0 framework showed better accuracy than many other CNNs. Therefore, we have used both distilBERT and EfficientNetB0 to achieve high accuracy. Authors [14] utilized distilBERT instead of RoBERTa for a late fusion strategy for a multi-modal fusion.

The foremost objective of the current research is to develop an end-to-end automated framework for fake news detection. Images and text are widely used data on social networking platforms; therefore, a multi-modal approach is presented that must have the ability to process

images and textual data. All in all, the primary goal is to develop an effective modal that should predict good results and address the limitations of the existing models.
The key contributions are highlighted as follows:

- First, we present an efficient multi-modal deep neural network for detecting fake news from social networking platforms. Our proposed modal utilizes EfficientNet-B0 and distilBERT, respectively, to process visual and textual data.
- Second, extensive computer simulations are conducted to determine the performance of the proposed modal. Results are collected and compared against the state-of-the-art methods. Real-world benchmark datasets such as MediaEval [15] (Twitter), Weibo, and Fakeddit (Reddit) are used during simulations.
- Third, results have been analyzed with regards to early vs. late fusion models and are reported as follows: (a) in the case of early fusion, learning intrinsic features of the fake news fusion of modalities have been analyzed; and (b) as far as late fusion is concerned, learning intrinsic features of each content type (visual or textual) have been analyzed at the individual channels and fusion is performed only at classification layer. This analysis reveals that late fusion showed a better response as compared to early fusion.

The rest of the paper is organized as follows: Section 2 presents related work; Section 3 discusses the proposed model; experimental setup, results, and discussion are given in Section 4; and Section 5 makes a concluding remark and highlights future research direction.

2. Related Work

Since fake news began demonstrating its harmful impacts, hence, fake news detection has been a significant research area. Some primary research was conducted using natural language processing (NLP) on textual data only [16]. The content of tweets or posts was studied for cosine similarity or analysis of sentiments to predict whether it is false or true. However, as there are many types of content in false news, a multi-modal methodology was introduced, in which a blend of two or three modalities was evaluated. In this section, we have presented a comprehensive literature review of existing work and highlighted its limitations. In subsections 2.1 and 2.2, we discussed literature based on single modularity, and multi-modularity approaches, whilst subsection 2.3 presents the limitations of the existing research.

2.1 Single modularity approach

The single component was initially used to predict fake or factual information on social platforms, such as using only images, only text, only context, and user profiles. Zhang and Ghorbani [17] shared a review of online fake news and covered various detection ways. However, multi-modal approaches were not covered. Mainly single modality techniques were discussed. Dwivedi et al. [18] conducted extensive research on online marketing and comprehensively discussed fake reviews and fake electronic words of mouth. Fake news was not covered. SRM-CNN based approach was initially recommended by Rao and Ni [19] to detect fake pictures. Some more mixed CNN models were suggested for fake image detection [20]–[22]. Jwa et al. [23] used a BERT transformer for fake news detection using news headlines on the text side. They first trained BERT with more corpus data from CNN to make it more specific to the news. Kaliyar et al. [24] improved on text-based modality. The model is adopted on a traditional text classification system that compromises an embedding layer as input where word embedding vectors are transferred. It utilizes GloVe to identify them as real or false as pre-trained for word embedding and unidirectional convolution neural networks for learning and training the document. The same authors Kaliyar et al. [25], experimented with

BERT and CNN models using text. The textual feature vectors generated by BERT were passed to the CNN model for feature learning. Another novel strategy applied by Kaliyar et al. [26] was to couple the tensor of textual with matrix–factorization and passed it through the CNN model. Working on the text itself, Goldani et al. [27] used three-layered CNN for fake detection. Kula et al. [28] used modified BERT to generate text embedding for fake news detection. They analyzed static and non-static word embeddings and trained over the CNN network for prediction. Text pattern mining over Twitter data was done by Diaz-G et al. [29]. They searched various text patterns in the self-collected tweets data of 6500 posts. Some innovative methods were also proposed which did not utilize NLP or CNN traditionally. One such idea was presented using an active method by Chen et al. [30]. The solution utilizes blockchain to identify the veracity of the fake content. Another novel approach using adversarial learning was employed by Wu et al. [31]. Here the standard features of the textual feature set were extracted using orthogonality constraints and KL-divergence. Apuke and Omar [32] studied Covid19 and fake news over social media users using a gratification framework extended by 'altruism' motivation. Working over traditional digital journalism where long articles of text are considered, Bonet-J et al. [33] proposed 5W1H based model, which is essential in lead construction. 5W's are What, Who, Where, When, Why, and How. Ghanem et al. [34] recommended leveraging suspect accounts' semantic and stylistic characteristics to determine the false integrity of news published by these accounts. Whereas, in [35], the social graphs methodology is used to detect hoaxes over social platforms. Sharma et al. [61] analyzed the sarcastic tweets and built a hybrid model to detect the sarcastic tweets. Eliciting out sarcastic tweets helps to improve the fake text accuracy as sometimes sarcastic tweets are marked as fake. Singh and Sharma [62] also created a single modal approach for fake image detection. The model could not only predict whether the image was fake or not but also highlighted the fake area within the fake image. They used Local Interpretable Model-agnostic Explanations (LIME) to highlight the fake part.

2.2 Multi-modal approach

More than one component of false news was examined in a recent study that used various modalities. Jin et al. [36] joined many content types and proposed a framework using a "Recurrent Neural Network" (RNN), having an attention mechanism for combining the visual, text, and social context features. Social context and text were initially joined with the LSTM network for fused vectors. The image vectors extracted from deep CNN were then linked with the resulting feature set. Sharma & Sharma [9] utilized the Cosine Similarity Index to identify false news by comparing text over pictures and headline text. The CNN-LSTM framework was utilized in the model. In [37]. web harvesting is used for data collection and applied reverse image search for the fake images and the text over fake images. "Event Adversarial Neural Networks" (EANN) was suggested by Wang et al. [38] to identify false news, extract event-invariant features, and aid in the fake detection on newly emerged events. The multimodal feature extraction is the first module in the design, followed by the false news detector and the event discriminator. The multimodal feature extractor's main task is to extract visual and textual characteristics from postings. Event discriminators are responsible for removing event-specific information while preserving event invariant properties across events. Cui et al. [39] developed Sentiment-Aware Multi-modal Embedding (SAME), a unique technique for identifying fabricated news that combines users concealed ideas from their posts into a unified deep multimodal embedding framework. Different networks manage diverse data first, such as text content, images, user profiles, and publishers. The adversarial technique is used in the following phase to discover semantically meaningful spaces for each data

modality. In the last step, the model describes a unique regularization loss that brings embeddings of important pairings closer together. An end-to-end solution was proposed by Khattaret al. [40] using the autoencoder Multimodal Variational Autoencoder (MVAE). The primary idea was to shape an autoencoder model. There are three primary modules in the proposed model: the encoder, the decoder, and the classifier module. In the encoder section of the model, there are two streams: text and visual, which are taught their respective functions. To generate text features, it uses Bi-directional LSTM, and for image features, it employs VGG19. Zeng et al. [41] proposed a model using an autoencoder and learning the correlation between the text and images for fake news detection. Zhou et al. [42] proposed the Similarity Aware Fake (SAFE) framework. The model separately calculates the likelihood of false news through text and visual learning. Later, along with the measured resemblance index between the image and text content, it considers all of these probabilities to mark it as false eventually or not. SpotFake is another well-known multi-modal framework that has been presented by Singhal [43]. Singh et al. [44] used a feature-based multi-modal approach that employed content, organization, emotion and manipulation features of fake news text and images. Giachanou et al. [45] employed a combination of text, visuals, and sentiments to detect fake news. They observed different datasets, and various combinations provided the optimum results. They later updated it with a more optimum framework [46]. Multi-modal Knowledge-aware Event Memory Network (MKEMN) proposed by Zhang et al. [47] to detect false news using the event-level model. It utilizes visual data and external information to support fake news detection. The method employed an event memory network to obtain event invariant vectors. Zhang et al. [48] proposed BDANN, a BERT based multi-modal using BERT for text analysis and VGG19 for images. Another explicit multi-modal was presented by Song et al. [49] and called “Crossmodal Attention Residual and Multichannel Neural Networks” (CARMN). This works by using an attention mechanism over cross-modalities. Merging user content features and news content features from Facebook was employed by Sahoo and Gupta [50]. They utilized machine learning and deep learning techniques to train these features and predict the news as false or genuine. Freire et al. [51] used crowd signals inspired by the meta information for detecting fake news. Raj and Meel [52] also utilized covNet for text and images but used the early fusion technique. Similarly, Wang et al. [53] also employed CNN with an attention mechanism. Madhusudhan et al. [54] utilized a similar approach, but they used SBERT and ResNet-18. Kang, Hwang, and Yu [55] used MCE, which required co-relation between modalities. Agarwal and Jalal [58] employed an ensemble neural network. Pradhan [59] did the sentiment analysis for fake detection but only on the text data on the YouTube comments.

2.3 Limitations of existing methods

As fake news consists of multiple contents, it is suggested to utilize a multi-modal approach for better results [11], [43]. Social media platforms do not readily share much information about user profiles, publishers, and network propagation; thus, a model utilizing text and images seems more practical. However, there are several downsides to the models discussed above, which use text and images. First, they have poor accuracy in real-world datasets [40][43][38] [57][60]. Second, they employ sub-activities such as learning cross-modal correlations or sub-tasks such as event discriminator and domain classifier [38], [40], [41], [47], [55]. Third, most models utilize early fusion techniques where multiple modalities are fused early, and fused features are learned [14], [43], [47]. In early fusion, the noise gets created as many feature vectors generated from the fused feature set will have an insignificant role in prediction.

Therefore, we have attempted to resolve these drawbacks. Our model utilizes optimized and latest models for text and image learning, which have better results on news datasets [12], [13]. For detection, our suggested model does not require any additional sub-activities. Each channel learns latent features at the individual level to mitigate noise from multi modalities fusion, and late fusion is done at the classification layer. The proposed modal produces more accurate results as compared to other state-of-the-art models. The proposed modal can be used directly by various fact-checking companies. Analyzing fake news necessitates a significant amount of physical labor and cost. It can recognize fake news automatically and is capable of handling both images and text. Another practical application is creating a plugin over browsers. Thus, when people browse through these social networking websites, they can be notified by the plugin as fake or real.

3. Proposed Scheme

Using a multi-modal approach, we suggest a practical approach to fix the issue of false news detection. The majority of fake news shared around social networking sites has both text and images. Automated feature extraction is the most crucial benefit of using a deep CNN. Thus, text and image modalities from the social networking sites are collected by the proposed model and are distributed for their respective feature extraction streams. We can address instances when the images are genuine but out of context by utilizing textual vectors. In these types of false content and false context cases, the text supports detecting fake news. **Fig. 1** and **Fig. 2**, respectively, display the proposed system design for late fusion and flowchart. Here we get the text embedding, image embedding, and feature vectors learned in their respective channels. Later, they are fused to generate the final vector for classification.

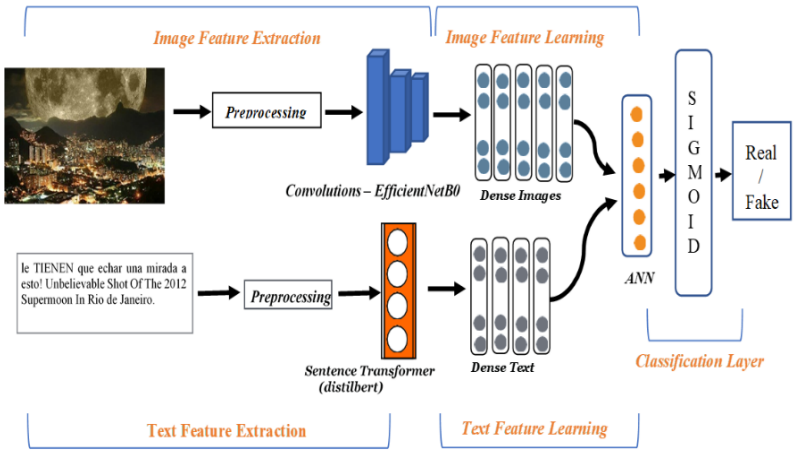

Fig. 1. Proposed system design - Late Fusion

The system architecture of the proposed model is shown in **Fig. 1**. **Fig. 1** is a more illustrative diagram of the flowchart. The architecture design is the result of iterative experiments done using the optimization library Talos. The final architecture is the optimized version after various structures and hyperparameters suggested by the Talos tool. There are three components:

a) Learning Textual Features vectors – The latent text feature is extracted and learned by this layer.

b) Learning Image Features vectors – It takes the hidden features of the images and learns them.
c) Classifier – Sigmoid is used for classification and labels them using the learning from integrated features.

List of symbols with their name used in manuscript.

FS_i - Feature sets of images
FS_t - Feature sets of text
FS_k - Fused Feature sets of Text and Image
$\emptyset$ - Activation Function
W_{tf} - Weights of text features
W_{if} - Weights of image features
$\hat{y}$ - Predicted probability
N - Numbers of items in the dataset

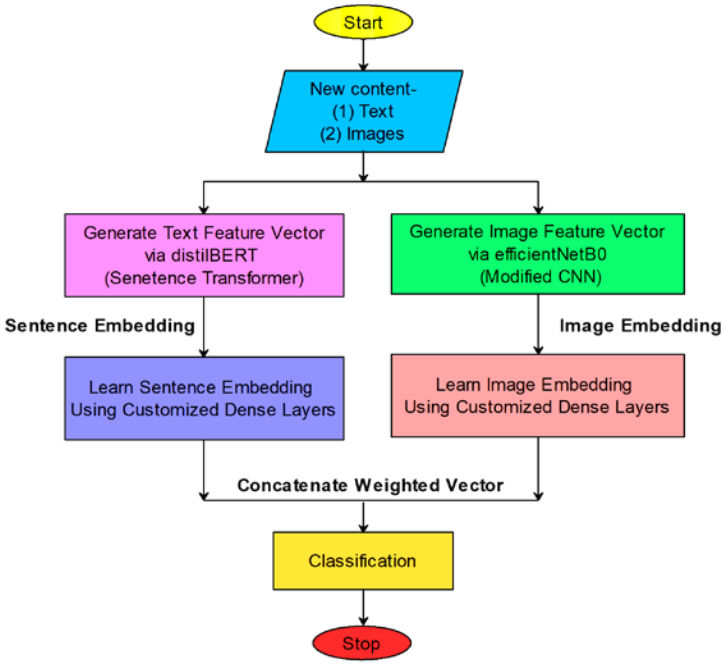

Fig. 2. Flowchart of the proposed framework

The suggested model uses the latest lightweight CNN model named EfficientNet to extract the latent characteristics of the images. The pictures are sent to the EfficientNetB0 model, and by utilizing transfer learning, image embeddings are generated from the third final layer of the pre-trained model. The image embeddings are passed to the stack of five completely linked dense layers to learn the image feature vectors. The image features are denoted as FS_i. Similarly, to generate text embeddings, the text is passed on to the sentence transformer distilBERT. Then, the textual embeddings are passed to four fully connected dense layers. Here the inherent features of the text are learned. The text features are represented as FS_t. The text and image feature vectors are then concatenated and transferred to the final classification layer. The Sigmoid predicts the odds of news being fake or real. The representation of model learning is as below.

Suppose there are "N" pairs to train, then modal $M = \{FS_k , G_k\}^N{}_{k=1}$. The FS_k features image and text vectors, and "G" is the truth label on the data mainly because it is a multimodal framework, elements from both modes are incorporated, as shown in equation (1).

$$FS_k = FS_t + FS_i \quad (1)$$

3.1 Learning textual features vectors

Textual content is initially preprocessed, where "Natural Language Processing" (NLP) methods are used to remove stopwords, punctuations. and other symbols. As there were tweets in other languages, they were translated to English using google translate API. The model uses the sentence transformer distilBERT for textual feature extraction to capture the latent semantic and contextual meanings. distilBERT is based on BERT.

BERT is a self-supervised pretraining approach developed and presented by Google that learns to predict purposely hidden (masked) text parts. BERT is based on the architecture of transformers. In BERT-base architecture, there are 12 layers of Encoders that are stacked together. It generates the embedding of 768 dimensions. It has two major parts in learning– Masked Language Modelling (MLM) and Next Sentence Prediction (NPS).

The authors fined-tuned BERT-base, made it lighter, and distilled it into distilBERT. In this paper, we have used the distilBERT-base uncased version model. distilBERT differs from BERT in majorly two ways. First, there are no token-type embeddings. This means there is no need to indicate which token belongs to which segment. Secondly, there is no pooler, meaning it works without NPS. In the comparison study, the creators of distilBERT [12] show that distilBERT performs better on specific datasets, especially on MRPC and SST-2. MRPC is extracted from online news sources. SST-2 dataset consists of sentences from online movie reviews. Both these datasets share similarities to our problem statement of fake news over microblogging sites. Thus, distilBERT is adopted in our proposed framework. distilBERT generates sentence embedding feature vectors which are passed to the four fully connected stacks of dense layers.

The learning of textual features may be modeled as presented in equation (2).

$$FS_t = \emptyset\ (W_{tf}\ FSt_{st}) \quad (2)$$

Where, $\emptyset$, W_{tf} and FSt_{st} respectively denotes activation function, weights from the last dense layers, and feature vectors generated from the transformer layer.

3.2 Learning image features vectors

The deep CNN has been utilized in image classification problems. Deep neural networks can saturate, and their accuracy is comparable to shallow neural networks. Therefore the compounding scaling formula for DNNwas given by EfficientNet from Google Brain. They also verified their multiple EfficientNet frameworks from EfficientNet-B0 to EfficientNet-B7. They demonstrated that their basic EfficientNetB0 is more effective than other 'deep neural networks, namely ResNet152, NASNet-A, and Inception-ResNetV2 [13]. EfficientNet-B0 outperforms many well-known "deep neural networks" with improved accuracy, working with fewer parameters and "floating-point operations per second"(FLOPS). The Key framework of the EfficientNetB0 is presented as follows:

- Swish Activation: This function is an accumulation of a linear and a sigmoid function. Ramchandran [56] showed that the Swish activation function had been shown to equal or surpass the "Rectified Linear Unit" (ReLU), particularly in image classification. The Swish has several benefits, including being bounded below and unbounded above and being non-monotonic. These characteristics enable it to beat ReLU in deep neural networks and evade dead neurons in the DNNs.

- Inverted Residual Block (MBConv Block): This creates a crosscut between the start and end of a convolutional block. A typical residual block with many channels has a wide -> narrow -> wide layout. There are a wide number of channels on the input layer that are compressed using a 1x1 convolution. For a 1x1 convolution, there is an increase in the number of channels so that input and output can be applied. In comparison, a narrow -> wide -> narrow approach follows an Inverted Residual Block, thus the inversion. The initial layer first broadens using a 1x1 convolution, after that employs a 3x3 depthwise convolution. Then, to add input and output, a 1x1 convolution is utilized to reduce channels.
- Squeeze-Excitation Blocks: This technique allocates weight to each channel rather than considering them all equally.

The full architecture of EfficientNetB0 is shown in **Fig. 3**. As per the formulae indicated, the other variations, such as B1, B2, B3, B4, B5, and B7, have identical structures but scaling specifically as per the formulae.

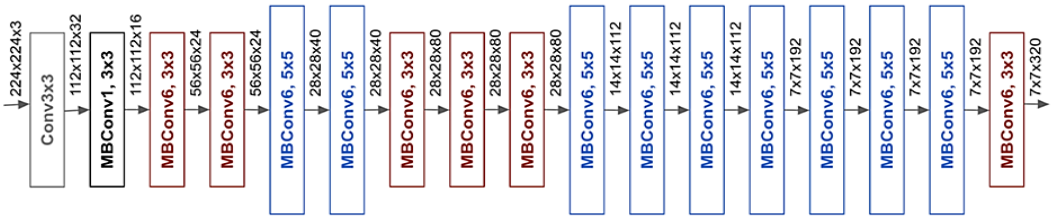

Fig. 3. As described by the authors, the architecture for the basic network EfficientNet-B0 [13]

EfficientNet-B0 is employed in the proposed model. The image embeddings are extracted from the third last layer of the pre-trained EfficientNet-B0. This layer has vectors for image attributes. The intrinsic features vector of images can be displayed using equation (3).

$$FS_i = \emptyset\ (W_{if} FSi_{effb0}) \tag{3}$$

Where, $\emptyset$, W_{if} and FSi_{effb0} respectively represent the activation function, the weight of vectors obtained from the last layer of the EfficientNet-B0, and output from the previous layer.

3.3 Classification layer

Before classification, the feature vectors generated from the text and image dense layers must be fused. The two different vectors set, FS_t*FS_i are joined into a vector of 2p dimensionality; this can be signified as $FS_k \in FS^{2p}$. Furthermore, we may designate the multimodal feature extractor as FE (IP; Θ_{fe}), where IP denotes the vectorized input data, and Θ_{fe} denotes the extractor's collection of parameters. The total mapping function can be represented as FE. After concatenating both modalities, the final feature set can be represented using equation (4).

$$IP_k = FS_t * FS_i$$
$$FS_k = FE\ (IP;\ \emptyset_{fe}) \tag{4}$$

After this, there is just one neural network layer before passing it on to the classifier. Tanh is the activation function utilized in dense layers. The experiment was also conducted using ReLU instead of tanh. Tanh performed better. The dense layer's vectors are sent to the sigmoid phase for categorization. We represent the predictor of the fake news from Sigmoid as OP (FS_k; Θ_{op}). The parameter's set of predictors is denoted by Θ_{op}, while OP denotes the mapping function. To increase learning, the Adam optimizer is used. The output generated from the predictor $\hat{y}$ for the multimodal event IP^j denotes the odds of the event. It is represented in equation (5).

$$\hat{y} = OP(FE(IP^j\ ;\ \emptyset_{fe});\ \emptyset_{op}) \tag{5}$$

Binary-cross entropy is used to determine the learning loss. The loss in binary-cross entropy is computed using equation (6).

$$\begin{aligned} \text{Loss}_{op}(\emptyset_{fe}, \emptyset_{op}) &= \sum_{i=1}^{n} y`_{i1} \log y_{i1} + y`_{i2} \log y_{i2} + \cdots + y`_{im} \log y_{im} \\ &= \partial \text{loss} / \partial y_{in} \\ &= \sum_{i=1}^{n} Y'_{im} / Y_{im} \\ &= \sum_{i=1}^{n} Y'_{i2} / Y_{i2} \end{aligned} \quad (6)$$

For the optimization of parameters $\emptyset_{fe}$ and $\emptyset_{op}$, there is a need to minimize the cross-entropy classification loss, which is denoted by equation (7).

$$(\emptyset^*_{fe}, \emptyset^*_{op}) = \min_{\emptyset fe, \emptyset op} \text{Loss}_{op} \quad (7)$$

Working of Algorithm:

Step 1. Provide the textual and visual (image) data to the model
Step 2. Textual data is pushed to textual channel and Image data is forwarded to image channel
Step 3. At both channels first preprocess the data. In textual data remove all stop words and perform stemming. In image channel, resize the image to 300 x 300 size.
Step 4. On textual side, extract text embeddings/vectors using DistilBERT transfer learning. On image side, extract the visual embeddings using EfficientNetB0 transfer learning.
Step 5. We get individual features sets/embeddings for text and image in respective channels
Step 6. Learn the feature learning of the text using four dense layers neural network. Here the weights at the nodes/features of dense layers are adjusted to correctly predict. Similarly, on the image channel, the image features/embeddings are learned using five dense layer neural network.
Step 7. Concatenate/Merge features set/weights from text & image using a single neural network layer
Step 8. Pass these fused feature sets to sigmoid for final prediction
Step 9. Check the predicted class probability with the actual class and calculate the error.
Step 10. Repeat this process (step 1- step 8) with all the datasets items and adjust the weights at the dense layers and at the artificial neural network (ANN) to minimize the error in prediction
Step 11. Stop the model learning if there is no significant improvement in the error reduction

Algorithm 1 presents the step-by-step working on the proposed modal. The IS_k represents the input set, FS_k denotes the feature set. FS_t denotes vectors from text and FS_i represents vectors of images.

Algorithm-1: Proposed Multi-Modal

Input: $IS_k = \{(TI_k, IgI_k)\}_{k=1}^{n}$
Output: $\hat{y}, \text{Loss}_{pr}, FS_k$

1. Set random values to W_{tf} and W_{if}
2. While max accuracy do
3. For each (TI_k)) do
4. Set values for FStk=φ ($W_{tfk}FSt_{db}$)
5. validate $\hat{y}$
6. End For
7. For each (IgI_k) do
8. Set values for FSik= φ ($W_{ifk}FSi_{effb0}$)
9. validate $\hat{y}$
10. End For

11. $FS_k = FS_{tk} \times FS_{ik}$
12. End While
12. Update FS_k as per equation (4)
13. Minimize $Loss_{pr}$ as per equation (6)
14. End

To customize multiple dense layers and to structure it so that it brings in optimal results, the following customization is added:

- **Dense layer-** In a neural network, a dense layer is a typical layer of neurons. Each neuron in the preceding layer receives information from all the neurons in the layer above it, making it highly linked. A weighted matrix, a bias, and the activations of the preceding layer make up this layer. Our suggested model has taken five dense layers for image learning and four dense layers for text feature learning.
- **Dropout-** A dropout is a regularisation approach in which a set of neurons is disregarded at random during training. On the forward pass, its significant contribution to downstream neuron activation is eliminated temporally. None of the weight changes is transferred to the neuron while on a backward pass. Dropout has been applied to the network's dense layers.
- **Flatten layer- Before** passing the feature vectors to dense layers, the output from the efficientNet-B0 and distilBERT must be brought into a singular matrix. Similarly, for classification at the concatenation phase, the feature vectors are flattened to a singular matrix.
- **Activation Function-** In a neural network, an activation function specifies how the weighted sum of the input is converted into an output from a node or nodes in a layer. The Rectifier Unit, or ReLU, is the most widely used activation function for CNN neurons' outputs. But we have used tanh as the activation function between the dense layers. We have experimented with ReLU also, but we got better results with tanh.

4. Experimental setup, results, and analysis

To ensure the model's effectiveness, we conducted an extensive experiment over the three publicly available social networking datasets, MediaEval [15], the Chinese dataset Weibo [36]. and Fakeddit [11]. The textual data of Weibo is in the Chinese language. As we consider both text and images as input, we have only considered posts associated with images from both datasets. We noticed that tweets were written in various languages, so we used the Google Translate library to translate them. We discovered that certain tweets had translation issues; therefore, those few tweets were left out of the experiment. Some of the postings were too lengthy for the Weibo dataset's sentence transformers; thus, they were ignored. Five tweets from the MediaEval dataset and 55 posts from Weibo datasets were ignored in this exercise because they did not fit into the architecture. These ignored tweets do not impact anything on the results, as for each image, there are multiple tweets/posts available.

4.1 Experimental setup

All images were resized to 300 X 300 size. Over the Google TensorFlow platform, the model was constructed using the Keras library on a machine with 32 GB RAM and having GPU Nvidia Quadro RTX 4000 8GBG DDR6. With a batch size of 128, the optimal findings were obtained in 200 epochs. Adam optimizer was used with a learning rate of $8e^{-4}$. We verified this architecture with tanh and ReLU activation functions. We got better results with tanh function as with ReLU the model was overfitting. Multiple iterations were required to select the right

combination of hyperparameters, employing different dropout probabilities and different batch sizes to get the right hyperparameter values. The number of potential permutations is lowered with each iteration dependent on the preceding iteration's performance. The Talos library is used to execute a random search and evaluate parameters in a random search. The Talos library was created to automate deep learning network hyperparameter tweaking and model assessment. The Talos library is open to public library for hyperparameter tuning for Keras. The library could be installed using pip (https://pypi.org/project/talos/)

The dataset was randomly divided into three groups, 75% to train, 10% to test, and 15% to validate. When the maximum level of accuracy was achieved, the final findings were recorded. An accuracy metric was employed to stop the network. **Table 1** presents the hyperparameters used in the proposed modal. **Fig. 4** shows the plot diagram, which demonstrates the implementation architecture.

Table 1. Hyperparameters values

S.No	Hyperparameter	Values
1	Dense layers	5
2	Flatten layers	1
3	Dropout layers	1 dropout layer with **0.4** value
4	Loss function	Binary-cross entropy
5	Activation function	Tanh (ReLU gave little lower results)
6	Learning rate	0.00008
7	Optimizer	Adam Optimizer
8	Epochs count	200
9	Batch size	128

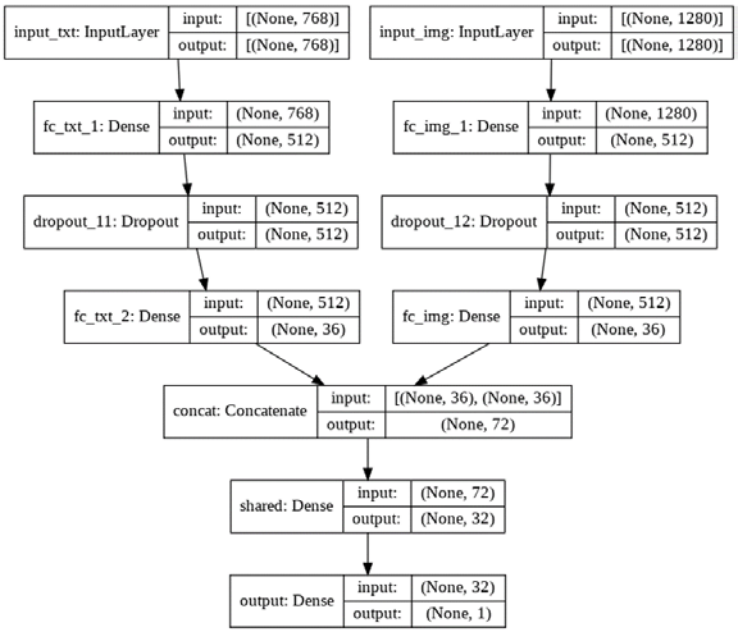

Fig. 4. Demonstration of implementation architecture of the proposed modal

4.2 Datasets

4.2.1 MediaEval

The social media dataset has 193 cases of real images, 218 cases of fake images, and 2 cases of altered videos. Around 6000 rumours and 5000 non-rumor posts from 11 events are included in the test collection. Around 2000 tweets of any type are included in the test sample. All the images and posts are collected from the real-world social networking application Twitter in the year 2016.

4.2.2 Weibo

The Weibo dataset [36] consists of data obtained from the Chinese real-world social platform Weibo, an authoritative news outlet by the Xinhua News Agency. It consists of false photographs and text retrieved from Weibo from 2012 to 2016. The dataset is reviewed by Weibo's official method of exposing rumors. The framework allows daily users to flag questionable Weibo tweets, which are then examined as fake or genuine by a reputed committee categorizing suspected messages. It has 3774 real images and 1363 fake images, along with the posts associated with them.

4.2.3 Fakeddit

The Fakeddit [11] is the latest and largest multi-modal dataset from the real-world social networking website Reddit. It contains over 1 million fake textual news data and over four hundred thousand multi-modal samples. The multi-modal samples have text and images. It has both 2-way and 6-way labeling. 2-way labeling is Fake and Real. As Reddit collects data from micro-sites like Twitter, Facebook, Instagram, and WhatsApp, this dataset has the largest diversified dataset.

4.2.4 Twitter Indian Dataset v3 [14]

A new dataset with an Indian focus has been produced to examine shifting trends on the microblogging site Twitter. We've compiled a list of fake and accurate news stories that were shared on the Twitter app. Posts and images covered are primarily from politics, Bollywood, and religion, as these are the frequently targeted areas in India for false news. The data was examined in two stages. To begin, all of the gathered information is double-checked using numerous well-known fact-checking websites operating in India, including alt news, India Today, and BOOMLive. Peer reviewers also do manual annotations in the second stage. The manual reviewers examined the news by logging on to the Twitter network and checking it. Only after two rounds of review labeling have been done. Also, it has been verified that there are no duplicates. The dataset has a total of 210 such images and collections of 1700 tweets. One hundred ten of the images are fake. Fake images are either morphed images or are out of context to the event. The dataset has events from November 2019 to June 2021, and they were all shared on Twitter in India.

4.3 Results and comparison

The primary purpose of this research was to extend and analyze the early fusion modal compared with late fusion. The text and image embeddings are concatenated early in the early-fusion framework, and fused vectors are learned over fully connected dense layers. This is the approach taken by some multi-modal frameworks like SpotFake [43], BDANN [48], and Singh & Sharma [14]. However, in late fusion, intrinsic features are learned at individual channels,

and then fusion occurs late at the last classification layer. It was observed that early fusion leads to noise information and hence impacts the accuracy of results. This noise is because, due to early- fusion, multiple feature vectors do not play a significant role in prediction. **Table 2**, **Table 3**, and **Table 4** document results for late and early. One can notice that the late fusion reveals better accuracy.

We compared the results with other benchmarked multi-modal techniques. MVAE [40], SpotFake [43], and BDANN [48] are among the benchmarking multi-modal models that have excellent accuracy over microblogging datasets. These techniques were proposed for both text and graphic information. MVAE employs a variational autoencoder component to learn the similarities between the two modalities, giving it an advantage over prior modals like EANN [38]. SpotFake and BDANN, on the contrary, employ the VGG19 image feature vector and the BERT text feature extraction transformer. Both feature learnings are fused in the early-fusion approach. Our proposed modal's accuracy surpasses other state-of-art models like MVAE, SpotFake, and BDANN.

We calculated Accuracy, Precision, Recall, F1 score, and AUC values to evaluate our methods' performance. Precision measures how to correct the model was in classifying positives. The recall is used to determine how many positives have been missed by the model. It is also called sensitivity or total positive rate (TPR). Accuracy is used to measure how accurately the model classifies data. F1-score determines the harmonic mean of precision and recall.

Table 2. Performance results of various models on a Twitter dataset.

Dataset	Model	Accuracy	Fake			Real		
			Precision	Recall	F1-Score	Precision	Recall	F1-Score
MediaEval (Twitter) [15]	VQA	0.631	0.765	0.509	0.611	0.55	0.794	0.650
	att-RNN	0.664	0.749	0.615	0.676	0.589	0.728	0.651
	EANN [38]	0.648	0.810	0.498	0.617	0.584	0.759	0.660
	SpotFake [43]	0.777	0.751	0.900	0.820	0.832	0.606	0.701
	MVAE [40]	0.745	0.801	0.719	0.758	0.689	0.777	0.730
	MKEMN [47]	0.664	0.753	0.537	0.627	0.611	0.805	0.695
	CARMN [49]	0.741	0.854	0.619	0.718	0.670	0.880	0.760
	BDANN [48]	0.830	0.810	0.630	0.710	0.830	0.930	0.880
	Early Fusion [14]	0.853	0.821	0.943	0.877	0.913	0.745	0.820
	Proposed – Late Fusion	0.864	0.840	0.930	0.88	0.900	0.780	0.840

Table 3. Performance results of various models on a Weibo dataset

Dataset	Model	Accuracy	Fake			Real		
			Precision	Recall	F1-Score	Precision	Recall	F1-Score
Weibo (Chinese) [36]	VQA	0.736	0.797	0.634	0.706	0.695	0.838	0.760
	att-RNN	0.772	0.797	0.713	0.692	0.684	0.84	0.754
	EANN [38]	0.795	0.827	0.697	0.756	0.752	0.863	0.804
	MVAE [40]	0.824	0.854	0.769	0.809	0.802	0.875	0.837
	MKEMN [47]	0.792	0.805	0.788	0.796	0.778	0.796	0.787
	CARMN [49]	0.853	0.891	0.814	0.851	0.818	0.894	0.854
	BDANN [48]	0.814	0.800	0.860	0.830	0.840	0.760	0.800
	Early Fusion [14]	0.812	0.851	0.784	0.816	0.744	0.826	0.782
	Late Fusion	0.814	0.803	0.863	0.836	0.841	0.767	0.808

Table 4. Performance results of various models on a Fakeddit dataset

Dataset	Model	Accuracy	Fake			Real		
			Precision	Recall	F1-Score	Precision	Recall	F1-Score
Fakeddit [11]	InferSent+ EfficientNet [11]	0.8339	-	-	-	-	-	-
	Proposed – Early Fusion	0.8620	0.83	0.79	0.80	0.89	0.86	0.87
	Proposed – Late Fusion	0.8880	0.85	0.87	0.86	0.92	0.90	0.91

Table 2 illustrates the results comparing various models over the MediaEval (Twitter) dataset to the proposed model. The proposed model results exceed the MVAE by 11.9%, SpotFake by 8.7%, and BDANN by 3.4%. The following factors are responsible for the higher results: First, the proposed model employs EfficientNet on image data, which works well compared to other image classification models (VGG-19, ResNet50). Inverted residual networks employed in EfficientNet perform better over images than VGG19 and ResNet50 [13]. Second, due to the application of late fusion of multi-modalities, less noise is generated. The model learns significant intrinsic features at the individual level, and only essential features are passed to the classification layer. The noise information is thus mitigated. On the textual side, distilBERT outperforms other Bidirectional LSTMs in MVAE to understand the context of brief tweets. Because tweets are short sentences with similar terms, word frequency is crucial in detecting them. Twitter tweets lack both. Thus, employing distilBERT supports better accuracy than BERT itself, as distilBERT showed more accuracy over online news datasets [12], [13]. **Fig. 5(a)**, **Fig. 5(b)**, and **Fig. 5(c)** show the confusion matrix and ROC curve obtained during the execution of these three publicly available datasets. **Table 3** compares the results of various models on the Weibo dataset to the proposed model. The accuracy of most models on the Weibo dataset is slightly higher than average, with a few exceptions. There are two explanations for the reduced accuracy. Due to the intricacy of the dataset, the translation is not particularly exact. Second, the posts in the Weibo dataset were longer than the shorter tweets on Twitter. **Table 4** compares the results of various models on the Fakeddit dataset to the proposed model. Our proposed framework accuracy is higher than the other models as late-fusion architecture works better than early fusion architecture.

4.3.1 Error analysis & limitations

Each channel learns its intrinsic feature vectors for that modality. For the image part, efficientNetB0 learns feature vectors with respect to color gradient changes besides the manipulated areas. The other feature vectors for image are edges, histogram of gradients, and sharp changes in the color maturity of the RGB channels. On the other hand, the distillBERT is a transformer. It learns the context of the statements. The feature vectors for its learning are the cosine distance of words with respect to fake and non-fake contexts. Though these feature vectors are robust enough to correctly classify the image and text as fake or not, there are still some edge cases described below where we observed failures.

There were a few takeaways from the falsely identified fake news. High-resolution images with only a tiny changed region were observed to be inadequately identified. We also saw cases where fake news had more irrelevant text posts than relevant posts and had correct predictions. These cases were more in Weibo datasets. Though the proposed framework can be used directly by the fact-checking industry, it still does not include satire news. It also does

not cover text present over images. Recently, videos have been shared more frequently. The framework does not consider fake videos. It is currently limited to text and images only.

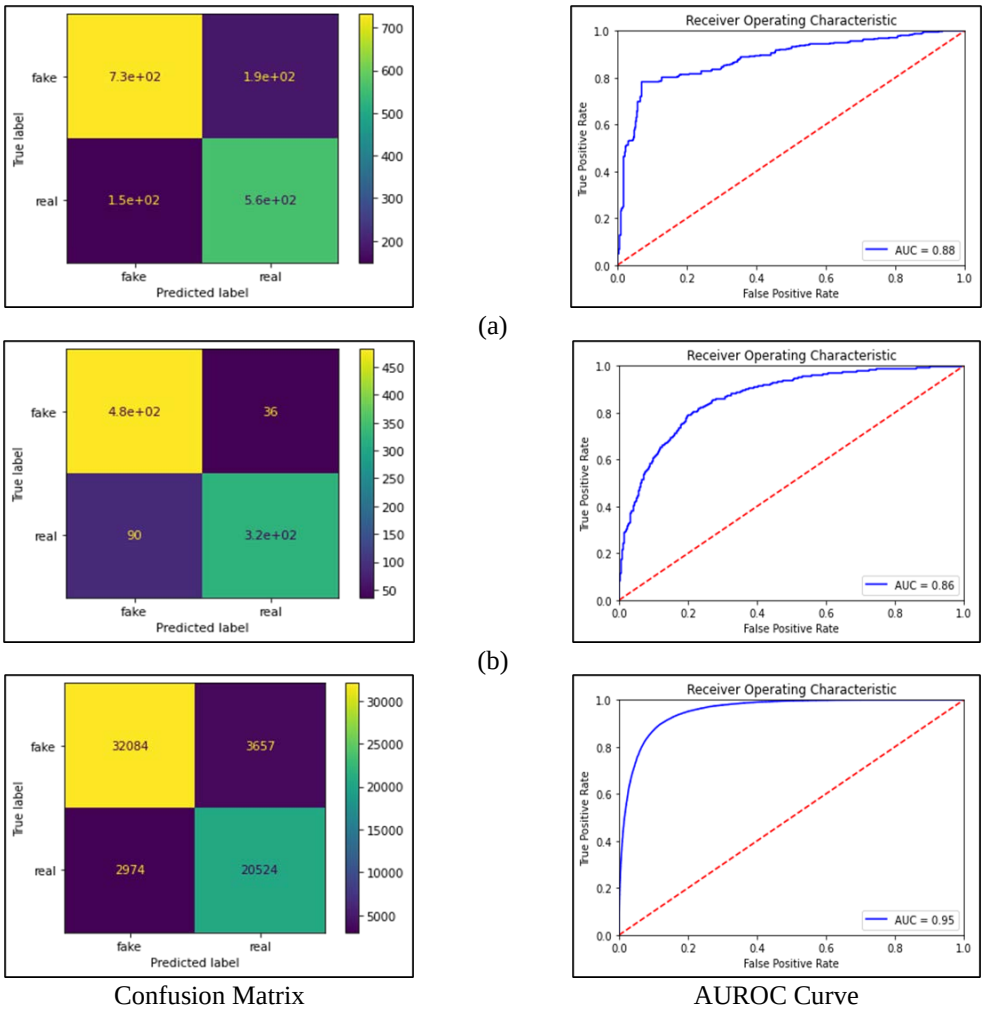

Fig. 5. Confusion Matrix, AUROC Curve respectively for (a) Twitter dataset; (b) Weibo dataset and (c) Fakeddit dataset

The confusion matrices in **Fig. 5** show that false-negative cases are more in Fakeddit datasets than on Twitter and Weibo. Fakeddit datasets have data from various platforms and not just social media platforms. Thus, a slightly higher number of false negatives are increased as the proposed model was trained on social media platforms specific datasets. If we closely look at the AUROC curve, we observe that the AUROC curve for the Twitter dataset is near the diagonal (TPR=FPR) as the model was trained over Twitter and denotes its slow learning rate. Once the model has been trained fully, the AUROC curve over the Weibo and Fakeddit dataset is smoother and closer to the top left part, indicating the model is more accurate.

4.3.2 Research over latest Indian dataset

MediaEval dataset and Weibo dataset can be considered outdated, as they pertain to particular incidents between 2012 and 2016. In the previous few years, there have been changes in the

way people use social networking sites. As a result, as part of our research, we generated a Twitter-based India viewpoint dataset. All of the news events in the dataset occurred between November 2019 and July 2021. This dataset contains 210 pictures and the tweets that go along with them. Of these, 110 pictures are fake, and the rest are genuine. India has multiple languages and has tweeted in regional languages. Only English-language tweets were taken into consideration. The majority of the news comes from the realms of politics, Bollywood, and religion.

We found several discrepancies between the newer Twitter dataset and the older Twitter dataset (MediaEval). There are three main explanations for these disparities. To begin with, changes to the Twitter platform rules for tweet length. Second, in India, people's attitudes toward social media platforms are changing. Third, the latest technological software is available for manipulations. The differences are as follows:

a. In India, people are writing more textual remarks than short tweets, as before. This is because the 140 character cap was expanded to 280 by Twitter in 2017. Such lengthy textual posts impair the learning of concise posts. This strengthens the thought mentioned in the article by [43]. Consider the following tweets as examples in **Table 5**.

Table 5. Examples of long tweets after 2017, when the length of the tweet was extended

S.No	Tweet	Words
1	“Ppl other than Assamese & other native ppl in Northeastern states had foreseen this many years ago but natives used to consider these pigs as their friends in endeavor to drive out other Non-Native Indians from those states. So, u ppl alone r responsible for this turmoil now”	276
2	“Whoever is supporting secularism will be treated as anti national and will be proved as traitor, culprit, ... now in India, This is the current trend in India..... Don't know why Indians have become enemy of india”	219
3	“Explain here The point was not to prove the pic is authentic or fake, the point was just by wearing skullcap you can't justify that they Muslims... Thanks for efforts by the way Bro.”	187

b. The second discrepancy observed was in politics-related images; more tweets were unrelated to the posted images/news. Twitter is presently being utilized as a grievance platform for political officials because of its broad reach. As a result, numerous tweets were unrelated to the image, as people expressed their dissatisfaction and grievances in posts rather than commenting about the image.
c. The third observation compared to the earlier dataset was that, in most images, only a small area of the image is manipulated, and a large part is genuine. Earlier, manipulated regions in images were significant, for example, in sandy hurricane images.

This illustrates that textual and visual cues have evolved with changing times. To improve accuracy and be in touch with the latest trends, models should be trained on the latest data. We also observed that style and content type differ from region to region. India's demographic’s style and images sharing content are different from the USA or European culture. This calls for the need for significant new region-specific datasets from social networking platforms. The dataset needs to be developed to keep up with the microblogging industry's evolving platform advancements. Region-specific and language-specific datasets should be created. Older datasets will not be compatible with today's social media network trends. Social Networking websites distribute more bogus news material than the rest of the internet, according to the CIGI-IPSOS poll [6]. The dataset need and research are, therefore, more for microblogging platforms than other news websites. When we validated our proposed framework over the

latest Indian dataset, we got an accuracy of 67.15% (**Fig. 6**).

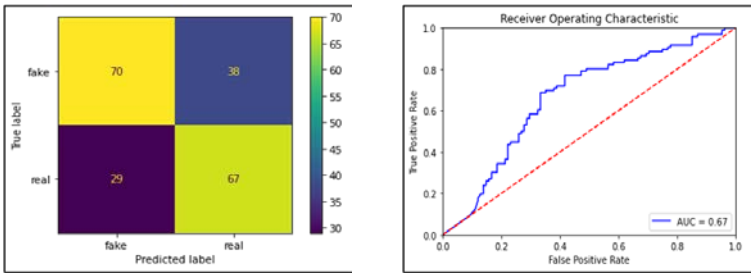

Fig. 6. Confusion Matrix and AUROC curve for the Indian v3 dataset

5. Conclusion

In this paper, a precise deep-learning-based multimodal approach has been proposed to detect fake news posted over social networking platforms. We discussed the limitations of the existing methods. Text and image modalities have been trained on individual streams and later fused. No additional sub-task was required to correlate the association between modalities. For mining the characteristics of text and image, the proposed modal employs EfficientNet-B0 and sentence transformer. The architecture was validated against popular microblogging platforms Twitter, Weibo, and Reddit. High accuracy of 86.48%, 82.50%, and 88.80% was achieved. This outcome exceeds other state-of-art multi-modal frameworks. A further experiment on late fusion against the early fusion of multi-modalities was also conducted. These models are capable of automatically marking the news as fake or real and, therefore, highly applicable in fact-checking industry. We have created the latest Twitter dataset to examine the most recent Twitter trends. The data was collected from the latest events of 2019 and 2020 from an Indian perspective. The differences were noticed, which indicated a dire need to create a dataset with the microblogging sites' latest data. This will keep the multi-modal models updated with changing trends in the microblogging industry. We noticed a few limitations of the proposed modal. There is no provision for detecting satire news. The text that is placed over the images is likewise ignored. Fake videos are also not covered, which have recently started to be shared more frequently. The suggested framework is not verifying fake news obtained from generative adversarial networks. These limitations will be taken up as a future research direction.

Acknowledgment

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A1B04032598).

References

[1] D. Chaffey, “Global social media research summary 2021,” 2021. [Online]. Available: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research (accessed Jan. 28, 2021).

[2] A. S. Neogi, K. A. Garg, R. K. Mishra, and Y. K. Dwivedi, “Sentiment analysis and classification of Indian farmers’ protest using twitter data,” *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, p. 100019, Nov. 2021. Article (CrossRef Link).

[3] J. Stoll, “Reading fake news about the coronavirus in Norway 2020,”. [Online]. Available: https://www.statista.com/statistics/1108710/reading-fake-news-about-the-coronavirus-in-norway-by-source/ (accessed May 02, 2020).
[4] D. Bunker, “Who do you trust? The digital destruction of shared situational awareness and the COVID-19 infodemic,” *Int. J. Inf. Manage.*, vol. 55, p. 102201, Dec. 2020. Article (CrossRef Link).
[5] L. Fazio, “Curbing fake news: Here’s why visuals are the most potent form of misinformation from,”. [Online]. Available: https://scroll.in/article/953395/curbing-fake-news-heres-why-visuals-are-the-most-potent-form-of-misinformation (accessed Feb. 01, 2021).
[6] “CIGI-Ipsos Global Survey on Internet Security and Trust,”. [Online]. Available: https://www.cigionline.org/internet-survey-2019 (accessed Jan. 28, 2021).
[7] K. Koc-Michalska, B. Bimber, D. Gomez, M. Jenkins, and S. Boulianne, “Public Beliefs about Falsehoods in News,” *Int. J. Press.*, vol. 25, no. 3, pp. 447–468, Jul. 2020. Article (CrossRef Link).
[8] N. McCarthy, “Report: Facebook Poses A Major Threat To Public Health,”. [Online]. Available: https://www.statista.com/chart/22660/health-misinformation-on-facebook/ (accessed Sep. 24, 2020).
[9] S. Sharma and D. K. Sharma, “Fake News Detection: A long way to go,” in *Proc. of 2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, pp. 816–821, Nov. 2019. Article (CrossRef Link).
[10] D. Varshney and D. K. Vishwakarma, “A review on rumour prediction and veracity assessment in online social network,” *Expert Syst. Appl.*, vol. 168, p. 114208, Apr. 2021. Article (CrossRef Link).
[11] K. Nakamura, S. Levy, and W. Y. Wang, “r/Fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection,” *arXiv preprint arXiv:1911.03854*, 2019. Article (CrossRef Link).
[12] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” *arXiv Prepr.*, 2019. Article (CrossRef Link).
[13] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *Proc. of in 36th International Conference on Machine Learning, ICML 2019*, vol. pp. 10691–10700, 2019. Article (CrossRef Link).
[14] B. Singh and D. K. Sharma, “Predicting image credibility in fake news over social media using multi-modal approach,” *Neural Comput. Appl.*, vol. 34, pp. 21503-21517, 2022. Article (CrossRef Link).
[15] C. Boididou, S. Papadopoulos, D.-T. D.-Nguyen, G. Boato, M. Riegler, S. E. Middleton, A. Petlund, and Y. Kompatsiaris, “Verifying Multimedia Use at MediaEval 2016,” in *Proc. of CEUR Workshop Proceedings*, vol. 1739, 2016.
[16] E. Saquete, D. Tomás, P. Moreda, P. Martínez-Barco, and M. Palomar, “Fighting post-truth using natural language processing: A review and open challenges,” *Expert Syst. Appl.*, vol. 141, p. 112943, Mar. 2020. Article (CrossRef Link).
[17] X. Zhang and A. A. Ghorbani, “An overview of online fake news: Characterization, detection, and discussion,” *Inf. Process. Manag.*, vol. 57, no. 2, p. 102025, Mar. 2020. Article (CrossRef Link).
[18] Y. K. Dwivedi et al., “Setting the future of digital and social media marketing research: Perspectives and research propositions,” *Int. J. Inf. Manage.*, vol. 59, p. 102168, Aug. 2021. Article (CrossRef Link).
[19] Y. Rao and J. Ni, “A deep learning approach to detection of splicing and copy-move forgeries in images,” in *Proc. of 2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1-6, 2016. Article (CrossRef Link).
[20] B. Bayar and M. C. Stamm, “A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer,” in *Proc. of the 4th ACM Workshop on Information Hiding and Multimedia Security - IH&MMSec ’16*, pp. 5–10, 2016. Article (CrossRef Link).

[21] Y. A. U. Rehman, L. M. Po, and M. Liu, "LiveNet: Improving features generalization for face liveness detection using convolution neural networks," *Expert Syst. Appl.*, vol. 108, pp. 159–169, Oct. 2018. Article (CrossRef Link).
[22] B. Liu and C.-M. Pun, "Exposing splicing forgery in realistic scenes using deep fusion network," *Inf. Sci. (Ny).*, vol. 526, pp. 133–150, Jul. 2020. Article (CrossRef Link).
[23] H. Jwa, D. Oh, K. Park, J. Kang, and H. Lim, "exBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder Representations from Transformers (BERT)," *Appl. Sci.*, vol. 9, no. 19, p. 4062, Sep. 2019. Article (CrossRef Link).
[24] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet – A deep convolutional neural network for fake news detection," *Cogn. Syst. Res.*, vol. 61, pp. 32–44, Jun. 2020. Article (CrossRef Link).
[25] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimed. Tools Appl.*, vol. 80, no. 8, pp. 11765–11788, Mar. 2021. Article (CrossRef Link).
[26] R. K. Kaliyar, A. Goswami, and P. Narang, "DeepFakE: improving fake news detection using tensor decomposition-based deep neural network," *J. Supercomput.*, vol. 77, no. 2, pp. 1015–1037, 2021. Article (CrossRef Link).
[27] M. H. Goldani, R. Safabakhsh, and S. Momtazi, "Convolutional neural network with margin loss for fake news detection," *Inf. Process. Manag.*, vol. 58, no. 1, p. 102418, Jan. 2021. Article (CrossRef Link).
[28] S. Kula, M. Choraś, and R. Kozik, "Application of the BERT-Based Architecture in Fake News Detection," in *Proc. of 13th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2020)*, pp. 239–249, 2021. Article (CrossRef Link).
[29] J. A. Diaz-Garcia, C. Fernandez-Basso, M. D. Ruiz, and M. J. Martin-Bautista, "Mining Text Patterns over Fake and Real Tweets," in *Proc. of IPMU 2020 Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pp. 648–660, 2020. Article (CrossRef Link).
[30] Q. Chen, G. Srivastava, R. M. Parizi, M. Aloqaily, and I. Al Ridhawi, "An incentive-aware blockchain-based solution for internet of fake media things," *Inf. Process. Manag.*, vol. 57, no. 6, p. 102370, Nov. 2020. Article (CrossRef Link).
[31] L. Wu, Y. Rao, A. Nazir, and H. Jin, "Discovering differential features: Adversarial learning for information credibility evaluation," *Inf. Sci. (Ny).*, vol. 516, pp. 453–473, Apr. 2020. Article (CrossRef Link).
[32] O. D. Apuke and B. Omar, "Fake news and COVID-19: modelling the predictors of fake news sharing among social media users," *Telemat. Informatics*, vol. 56, p. 101475, Jan. 2021. Article (CrossRef Link).
[33] A. Bonet-Jover, A. Piad-Morffis, E. Saquete, P. Martínez-Barco, and M. Ángel García-Cumbreras, "Exploiting discourse structure of traditional digital media to enhance automatic fake news detection," *Expert Syst. Appl.*, vol. 169, p. 114340, May 2021. Article (CrossRef Link).
[34] B. Ghanem, S. P. Ponzetto, and P. Rosso, "FacTweet: Profiling Fake News Twitter Accounts," in *Proc. of SLSP 2020 Statistical Language and Speech Processing*, pp. 35–45, 2020. Article (CrossRef Link).
[35] F. Tchakounté, K. Amadou Calvin, A. A. A. Ari, and D. J. Fotsa Mbogne, "A smart contract logic to reduce hoax propagation across social media," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3070–3078, Jun. 2022. Article (CrossRef Link).
[36] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs," in *Proc. of the 25th ACM international conference on Multimedia*, pp. 795–816, Oct. 2017. Article (CrossRef Link).
[37] D. K. Vishwakarma, D. Varshney, and A. Yadav, "Detection and veracity analysis of fake news via scrapping and authenticating the web search," *Cogn. Syst. Res.*, vol. 58, pp. 217–229, Dec. 2019. Article (CrossRef Link).

[38] Y. Wang et al., "EANN: Event adversarial neural networks for multi-modal fake news detection," in *Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 849–857, Jul. 2018. Article (CrossRef Link).
[39] L. Cui, S. Wang, and D. Lee, "Same: Sentiment-aware multi-modal embedding for detecting fake news," in *Proc. of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2019*, pp. 41–48, Aug. 2019. Article (CrossRef Link).
[40] D. Khattar, J. S. Goud, M. Gupta, and V. Varma, "MVAE: Multimodal Variational Autoencoder for Fake News Detection," in *Proc. of The World Wide Web Conference*, pp. 2915–2921, May 2019. Article (CrossRef Link).
[41] J. Zeng, Y. Zhang, and X. Ma, "Fake news detection for epidemic emergencies via deep correlations between text and images," *Sustain. Cities Soc.*, vol. 66, p. 102652, Mar. 2021. Article (CrossRef Link).
[42] X. Zhou, J. Wu, and R. Zafarani, "SAFE: Similarity-Aware Multi-modal Fake News Detection," in *Proc. of Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 354–367, 2020. Article (CrossRef Link).
[43] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, "SpotFake: A Multi-modal Framework for Fake News Detection," in *Proc. of 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, pp. 39–47, Sep. 2019. Article (CrossRef Link).
[44] V. K. Singh, I. Ghosh, and D. Sonagara, "Detecting fake news stories via multimodal analysis," *J. Assoc. Inf. Sci. Technol.*, vol. 72, no. 1, pp. 3–17, Jan. 2021. Article (CrossRef Link).
[45] A. Giachanou, G. Zhang, and P. Rosso, "Multimodal Fake News Detection with Textual, Visual and Semantic Information," in *Proc. of International Conference on Text, Speech, and Dialogue*, pp. 30-38, 2020. Article (CrossRef Link).
[46] A. Giachanou, G. Zhang, and P. Rosso, "Multimodal Multi-image Fake News Detection," in *Proc. of 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 647–654, Oct. 2020. Article (CrossRef Link).
[47] H. Zhang, Q. Fang, S. Qian, and C. Xu, "Multi-modal Knowledge-aware Event Memory Network for Social Media Rumor Detection," in *Proc. of the 27th ACM International Conference on Multimedia*, pp. 1942–1951, Oct. 2019. Article (CrossRef Link).
[48] T. Zhang et al., "BDANN: BERT-Based Domain Adaptation Neural Network for Multi-Modal Fake News Detection," in *Proc. of 2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Jul. 2020. Article (CrossRef Link).
[49] C. Song, N. Ning, Y. Zhang, and B. Wu, "A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks," *Inf. Process. Manag.*, vol. 58, no. 1, p. 102437, 2021. Article (CrossRef Link).
[50] S. R. Sahoo and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," *Appl. Soft Comput.*, vol. 100, p. 106983, Mar. 2021. Article (CrossRef Link).
[51] P. M. Souza Freire, F. R. Matias da Silva, and R. R. Goldschmidt, "Fake news detection based on explicit and implicit signals of a hybrid crowd: An approach inspired in meta-learning," *Expert Syst. Appl.*, vol. 183, p. 115414, Nov. 2021. Article (CrossRef Link).
[52] C. Raj and P. Meel, "ConvNet frameworks for multi-modal fake news detection," *Appl. Intell.*, vol. 51, no. 11, pp. 8132–8148, Nov. 2021. Article (CrossRef Link).
[53] Y. Wang, F. Ma, H. Wang, K. Jha, and J. Gao, "Multimodal Emergent Fake News Detection via Meta Neural Process Networks," in *Proc. of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3708–3716, Aug. 2021. Article (CrossRef Link).
[54] S. Madhusudhan, S. Mahurkar, and S. K. Nagarajan, "Attributional analysis of Multi-Modal Fake News Detection Models (Grand Challenge)," in *Proc. of 2020 IEEE 6th International Conference on Multimedia Big Data, BigMM 2020*, pp. 451–455, 2020. Article (CrossRef Link).
[55] S. Kang, J. Hwang, and H. Yu, "Multi-Modal Component Embedding for Fake News Detection," in *Proc. of 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, pp. 1–6, Jan. 2020. Article (CrossRef Link).

[56] P. Ramachandran, B. Zoph and Q. V Le, "Swish: A Self-Gated Activation Function," in *Proc. of 6th Int. Conf. Learn. Represent. ICLR 2018 - Work. Track Proc.*, no. 1, 2018.

[57] S. C. Agrawal and A. S. Jalal, "Distortion-free image dehazing by superpixels and ensemble neural network," *Vis Comput*, vol. 38, pp. 781–796, 2022. Article (CrossRef Link).

[58] R. Pradhan, "Extracting Sentiments from YouTube Comments," in *Proc. of the 2021 Sixth International Conference on Image Information Processing (ICIIP)*, Shimla, India, 26–28 November 2021. Article (CrossRef Link).

[59] R. Agarwal, A. S. Jalal, S. C. Agrawal, and K. V. Arya "Fake and Live Fingerprint Detection Using Local Diagonal Extrema Pattern and Local Phase Quantization," in *Proc. of Conference ICDLAIR2019*, pp. 73-81, 2019. Article (CrossRef Link).

[60] V. Jain, M. Agrawal, and A. Kumar, "Performance Analysis of Machine Learning Algorithms in Credit Cards Fraud Detection," in *Proc. of the 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, pp. 86–88, 4–5 June 2020. Article (CrossRef Link).

[61] D. K. Sharma, B. Singh, S. Agarwal, H. Kim, and R. Sharma, "Sarcasm Detection over Social Media Platforms Using Hybrid Auto-Encoder-Based Model," *Electronics*, 11, 2844, 2022. Article (CrossRef Link).

[62] B. Singh, D. K. Sharma, "SiteForge: Detecting and localizing forged images on microblogging platforms using deep convolutional neural network," *Computers & Industrial Engineering*, Vol. 162, 2021. Article (CrossRef Link).

Dilip Kumar Sharma is B.E. (CSE), M.Tech. (CSE) and Ph.D in Computer Engineering. He is a Senior Member of IEEE, ACM, and CSI. He is Fellow of IE and IETE. He is working as a Dean of Internatioanl Relations & Academic collaborations and Professor in the Department of Computer Engineering & Applications, GLA University, Mathura, India from March 27, 2003, to till now. He has edited 07 Books volumes and 01 is in the ongoing stage. He has published more than 190 research papers in International Journals /Conferences of repute indexed in SCI, Scopus, and DBLP databases and participated in 60 International/National conferences. He has published 06 patents.

Bhuvanesh Singh is MCA and PhD in computer applications from GLA University, India. He is pursuing his master of sciences at University of St. Thomas, Minnesota, U.S.A He is currently working in the field of data science at Ford Motors, the USA on telemetric data. He has published many research papers in International Journals /Conferences of repute indexed in SCI, Scopus indexing. He has attended many online courses/seminars/workshops related to technologies. His current research interests include artificial intelligence, machine learning and computer vision.

Saurabh Agarwal is an associate professor at Amity University, Noida, India, from 2018. He received his B.E. degree in Computer Science & Engineering from Barkatullah University, Bhopal, in 2003 and his M.Tech degree in Software Engineering from APJ AKTU, Lucknow in 2010. He received his Ph.D. in Computer Engineering from the University of Delhi in 2017, India. His current research interests are image forensics, computer vision, and artificial intelligence.

Hyunsung Kim received the M.Sc. and Ph.D. degrees in computer engineering from Kyungpook National University, Korea, in 1998 and 2002. He is a Professor in the Department of Cyber Security, Kyungil University, Korea since 2012. Furthermore, he is currently a visiting professor at the Department of Mathematical Sciences, Chancellor College, University of Malawi, Malawi since 2015. He also was a visiting researcher at Dublin City University in 2009. From 2000 to 2002, he worked as a senior researcher at Ditto Technology. His research interests include cryptography, VLSI, authentication technologies, network security, ubiquitous computing security, and security protocol.

Raj Sharma is studying in B. Tech Computer Science Engineering with specialization in Cloud Computing from Sri Ramaswamy Memorial Institute of Science and Technology, India. He has attended many online courses related to technologies. His current interests are artificial intelligence, internet of things, cloud web services.