International Journal of Advanced Technology and Engineering Exploration, Vol 11(114)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2023.10102314

Research Article

Multimodal fake news detection using hyperparameter-tuned BERT and ResNet110

Chetan Agrawal[1*], Anjana Pandey[2] and Sachin Goyal[2]
Research Scholar, Computer Science & Engineering, University Institute of Technology Rajiv Gandhi Proudyogiki Vishwavidyalay (UIT - RGPV), Bhopal, India[1]
Associate Professor, Information Technology, University Institute of Technology Rajiv Gandhi Proudyogiki Vishwavidyalay (UIT - RGPV), Bhopal, India[2]

Received: 11-September-2023; Revised: 21-May-2024; Accepted: 23-May-2024

Abstract
Globally, the usage of social media has significantly increased and has become the most common way for people to deplete news. The easy sharing of multimedia content on social media has caused the fake news dimension, which threatens the stability as well as security of the society. Fake news detection (FND) in social media becomes challenging, because of which various tools are developed to detect them. Multimodal FND aims to determine fake data by text as well as images. Most commonly, researchers identify fake news only as text, but not as images. This research proposes a multimodal method for the detection and classification of fake news into real or fake. The proposed multimodal-based convolutional neural network (CNN) combines the designs of both text and image of fake news. This method utilizes two classification methods named hyperparameter tuning based bidirectional encoder representations from transformers (HTBERT) for text, and ResNet110 for images. Fakeddit dataset has used to estimate and evaluate the performance.
The experimental results of the proposed ResNet110+HTBERT model achieves respective accuracy, precision, recall and F1-score values of about 0.931, 0.944, 0.942, and 0.946, which is superior when compared to the existing methods, recurrent CNN (RCNN) and fine-grained multimodal fusion network (FMFN). From the analysis, it is evident that the proposed method ResNet110+HTBERT achieves an accuracy of 0.931, and hence shows better results for overall metrics when compared to the existing methods of RCNN and FMFN.

Keywords
Convolutional neural network, Hyperparameter tuning based bidirectional encoder representations from transformers, Multimodal fake news detection, ResNet110, Social media.

1.Introduction
Recently, people obtain updated news from various platforms including online sources and social media sites [1]. The social media resources Facebook, Twitter and Sina Weibo have become the top sources of news as thousands of users read news through these social media platforms. Social media has aided users to acquire news, indicate perspectives and communicate personal judgements with others [2]. The huge evolution of web technology has made it possible for users to post both real and fake news on social media [3, 4]. The blogs, headlines and social media messages are deliberately put forward as ambiguous for various reasons [5]. Many communities progressively receive information and devour news through social media. It allows the community to share information, where their information is in the form of both text and images [6]. The news arises from various platforms, making it a challenging task to identify the credibility of the posts and news [7]. The information provided in images is stronger than on text [8]. Social networks provide the benefit of transmission of multimedia information easily at low costs with a faster distribution, which easily and rapidly benefits and attracts people [9]. A growing number of users have begun fabricating fake news on social media with the intent of luring people and communities [10]. Fake news is rapidly increasing, causing a negative impact and constituting a significant menace to the individuals, society, economy and health [11].

Automatic fake news detection (FND) has become a major concern as it is challenging to detect fake news

*Author for correspondence

manually. Hence, automatic FND has turned out to be a recent research subject [12]. The FND methods effectively determine false news and are supportive for an administrator to remove fake news from the social media [13]. However, FND provide a fine-grained multi-classification problem during implementation by developing a number of categories for the collection of data [14]. The FND supports readers in determining fake news and bias in news articles, and hence reduces the spread of fake news [15, 16]. The initial process of FND is to identify if the fake news is a text or image. For this, the multimodal approach is essential to provide correlative benefits for FND [17]. The features of multimodal are anticipated to the most favourable in FND compared to the existing models [18]. The existing researches suggest that the ensemble learning and machine learning (ML) algorithms train the exact high-level representations obtained from the postings on social media sources for recognition [19]. Few existing researchers have tried to physically develop a series of features that are provided to the ML for identifying fake news, however these methods still consume more time and have poor conception [20]. For network administrators, the manual rejection of fake news one by one is expensive and laborious. Additionally, these methods are enduring with their flexibility characteristic and solve the overfitting problem. To overcome these problems, a multimodal method is proposed for FND in the formats of both text and images. The proposed multimodal-based convolutional neural network (CNN) combines the design of text and image. This method utilizes two classification methods namely, hyperparameter tuning based bidirectional encoder representations from transformers (HTBERT) for text, and ResNet110 for images. The primary contributions of this research are discussed as follows:

- This research proposes multimodal HTBERT and ResNet110 architecture for FND to classify the news as real or fake.
- This method utilizes two classification methods namely, HTBERT for text and ResNet110 for image. The multimodal FND makes use of the Fakeddit dataset for the effective performance of the model.
- The algorithm of error level analysis (ELA) focuses on the malignant spliced and fake image attributes which is better than sending the image directly into the frequency domain by evaluation.

The rest of the paper is arranged as follows: Section 2 discusses the recent research on task assignment problems and section 3 provides the proposed work of this paper. The assessment results are discussed in section 4. The discussion and limitations of the proposed method are explained in section 5, while section 6 presents the overall summary and future work of this research.

2.Literature survey

Liu et al. [21] implemented a semantic gap bridging among text and image by utilizing the caption-based approach to capture semantic information from images. This model optimized the use of image by combining entity features with global features to enhance the accuracy. It also leveraged image caption technology to produce image data and integrate it into the original text, so as to bridge the gaps. This method provided better performance and significant improvement than the other methods. However, the model gave rise to insufficient data due to its expensive annotation.

Segura-bedmar and Alonso-bartolome [22] implemented a fake news fine-grained classification on the dataset of Fakeddit, utilizing both the approaches of single and multiple models. The multiple modals used an architecture of CNN which combined the data of image and text, whereas the single modal only utilized the text. The advantage of this modal was that it provided better accuracy, while the image information expanded the scope for a good FND performance; but still, aggressive training gave rise to time complexity.

Ying et al. [23] developed a novel method for end-to-end multimodal cross-attention networks with multiple stages of textual content. This method jointly integrated the duplicate relationships of text and visual data, as well as the variant modalities of social media news in a unified method. ResNet and bidirectional encoder representations from transformers (BERT) classification were pretrained for work, so as to produce effective likeness for the regions. This method provided better performance than the other methods by combining the multi-level models. However, the method was unable to ensure that the central hidden state captured ample textual linguistics.

Guo and Song [24] introduced multiple model FND with the attention and pooling methods of neural networks. This method initially utilized two multimodal learning stages, averaging pooling and multimodal fusion using the dot-product attention. This method utilized hidden knowledge merge by both temporal and spatial effects. The model provided better results than other methods by utilizing attention and pooling blocks, but was ineffective in ensuring a similar relation in attention.

Wang et al. [25] implemented a fine-grained multimodal fusion network (FMFN) as a nuanced method for the fusion of textual and visual features for effective FND. The suggested approach used scaled dot-product attention method to enhance both features and fuse the improved features, thereby capturing the province among features. This presented method showcased significant improvements in the detection of fake news, but it had a number of representations which were complex to distinguish.

Jing et al. [26] implemented a multimodal progressive fusion network (MPFN) for multiple FND. The MPFN minimized fake news with determinate baseline techniques by using image superficial information and deep information consideration. This method was most widely used for multimodal fused features and modalities interactions in FND. This method improved the modal's performance by utilizing the features of the modalities, but the manual labelling performance was not supportive for fake news identification in the dataset's results.

Kumari and Ekbal [27] implemented attention-based multimodal factorized bilinear pooling (AMFB). This framework utilized the post-textual and image data as input to determine whether that post was real or fake. An attention based multi-level CNN-recurrent neural network (ABM-CNN-RNN) was utilized for the extraction of features from an image. This method provided better feature extraction (FE) and fusion performance than the previous methods, but could not extract post-specific features from a difficult post. Rai et al. [28] developed an approach for fake news classification based on the titles of news with the use of content-based classification method. This approach deployed the hybrid model of BERT and long short-term memory (LSTM) for classification to classify the news as legal or fake. This method classified the news based on semantic attributes of news articles or reports including grammatical, semantical and synthetical perspectives. This method utilized FakeNews Net dataset consisting of two subsets, PolitiFact and GossipCop for training and validation. This method achieved better classification performance because of the LSTM having a greater capability to catch the semantic, as well as long-distance relationships. Nonetheless, this model was enormous due to corpus data, and its structure of training.

Li et al. [29] presented a new method of semantic-enhanced multimodal fusion network for FND. This method contained multiple subnetworks such as multimodal fusion, fake news detector, as well as variation network of an event domain. It employed CNN to fuse the multimodal data, and later embraced a domain variation network to learn the transferred attributes among events. This method also examined the distribution of language statistics of social media to traverse an optimal selection of pre-trained BERT. This method effectively captured the mutual attributes between events, and hence brought profit for FND.

Mehta et al. [30] introduced a natural language processing (NLP) framework of BERT for fake news classification. This method fine-tuned a BERT for specific domain datasets, aside from employing human justification and metadata for the extended performance of this method. Further, it analysed that the deep-contextualizing nature of BERT was efficient for this work, also acquiring an efficient improvement on the binary classification. It effectively classified 6 label classification models. Nonetheless, this method consumed more time for training.

Palani et al. [31] implemented the CapsNet and BERT (CB) for an automated detection of fake news. The CB-Fake model utilized both visual and textual data from social media news articles for estimating as them being real or fake. This method included BERT to extract the features of text which protected semantic relationships among the words. The CapsNet caught significant visual attributes from the image. Then, those attributes were integrated to acquire the abundant information that supported to examine if the news was real or fake, yet this model had high computational cost.

Kaliyar et al. [32] integrated various parallel blocks of single-layer deep CNN which had various kernel sizes and filters with BERT for efficient learning. This method was developed on bidirectional transformer encoder top based on BERT. It achieved better performance and did not need handwritten features, but demanded more computations due to its size. Aslam et al. [33] presented an ensemble-based deep learning (DL) approach for classification of fake news. This method utilized two DL methods namely, bi-directional LSTM (Bi-LSTM), and gated recurrent unit (GRU) for textual features, while for the balance features, dense DL methods were utilized. This method applied the NLP techniques on the statement features and utilized the LIAR dataset for classification whether the news was fake. This method achieved a superior performance, as compared to other methods. Nevertheless, the Bi-LSTM approach consumed more time to train the model than the traditional LSTM.

Choudhary and Arora [34] developed a linguistic feature-based driven DL approach for effective detection and classification of fake news. The linguistic approach was developed to identify the content effects which introduced language-driven features. This approach extracted the synthetic, sentimental, readability and grammatical aspects of particular news items. It achieved commendable outcomes by deploying neural-based sequential learning. Yet, the language driven model needed an operation to control the time-consumption and the handcrafted feature problems that could not control the imprecation of dimensionality problem.

Xue et al. [35] implemented a multimodal consistency neural network (MCNN) for identifying the text and images of fake news. This method employed ranch network for extracting the visual semantic features in visual perception of fake news. In visual tampering FE, the ELA and CNN were developed to examine the intellect and originality of the pictures given manually in fake news. The visual tampering FE module effectively detected the malicious tampered images of fake news. But this method needed huge labelled data for model training.
Guo [36] introduced multimodal FND using a mutual attention neural network (MANN) which learnt the relationship among different modalities on the Weibo dataset. The model comprised four phases: a multimodal FE, fusion, detection, and irrelevant event discrimination. But, the features on the VGG16 and VGG19 layers were tensors with dimensions (channels, width, height). Additionally, some challenges arose due to the varying sizes across different layers.

Yang et al. [37] developed the multimodal transformer for FND using thermogravimetric analysis (TGA) on Weibo dataset and Twitter dataset. TGA, a transformer-based multimodal approach, consisted of four key steps: a text and image FE, fusion and a classification. Yet, in the TGA, the mass loss of volatiles was not equivalent to the formation of degradants, significantly impeding its ability to provide consistent universal indicators of the actual extent of degradation.

Sastrawan et al. [38] introduced a CNN–RNN-based FND model which underwent testing and training on the datasets of Indian society of organ transplantation (ISOT), fake news dataset, fake or real news dataset, and FND dataset. To address data imbalances between classes, the datasets underwent augmentation process using the back-translation method. However, a challenge encountered was with the long sequences. In dealing with very long sequences, RNNs faced memory limitations, potentially impacting their ability to effectively process extensive information.

Al et al. [39] introduced an approach to improve the existing methods' performances through the utilization of ensemble DL-based on attention mechanisms. In this approach, the achievement of ensemble approach depended on various learners. The loss function enforced every learner to perform various news content parts and obtained better classification accuracy. Furthermore, the learners were developed on basic deep feature extractor but differed from attention approaches. The parameter numbers were efficiently minimized and overfitting was solved.

Uppada et al. [40] presented a framework that flagged the fake news by embedding visual data with text. The suggested framework performed on the data obtained from a benchmark dataset. This approach had various architectures to learn visual and linguistic approaches from individual news items. Various datasets were analysed while the features extracted from those perceptions were text-based supported data.

The major limitations of previous work include manual labeling performance, which is inadequate for fake news identification because it is time-consuming to train the model and increases computational complexity. To address these issues, a multimodal FND technique is proposed in this manuscript. The suggested method uses HTBERT and RsNet110 to classify news as real or fake across both text and image formats, achieving superior accuracy.

3.Methods

In this work, the proposed multimodal FND is fitted by giving text and image news. The model's goal is to determine if the news is real or fake. The proposed method is categorized into three main parts: textual features, visual extractor, as well as feature fusion of textual and visual data. *Figure 1* depicts the block diagram of the proposed multimodal FND.

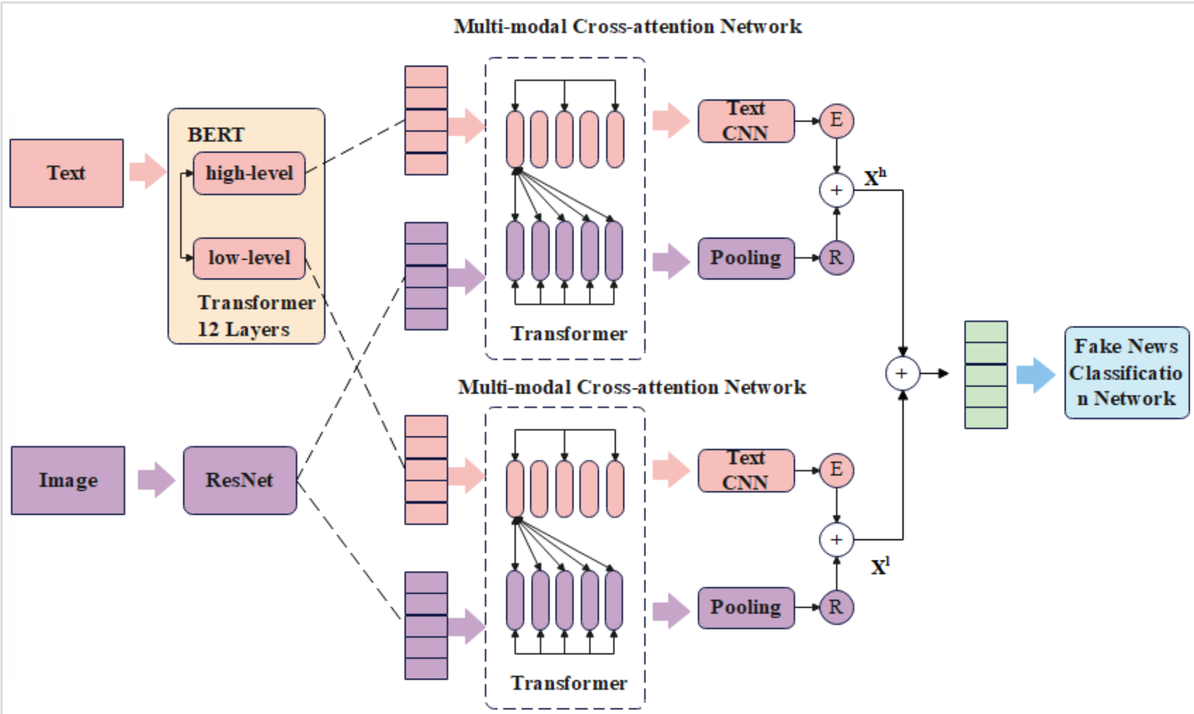

Figure 1 Proposed method for multimodal FND

3.1Dataset
Primarily, this research obtains the standard multimodal FND dataset of Fakeddit [41]. This dataset involves comments, text piece, related images, context features as well as ground truth labels. Then, the collected dataset is provided for the pre-processing step.

3.2Pre-processing
In this step, this research pre-processes the collected data of both text and image. Primarily, this process eliminates the instances which involve either text or image for achieving the better result in multimodal data. In textual data, the methods of tokenization, stemming and lemmatization are utilized to remove the punctuation marks, stop words etc. Tokenization is the procedure of dividing the text words into a list of tokens which are further utilized for lexical analysis, where the data fields are changed into tokens. For the image data, the normalization technique is used for converting raw data into a suitable format as the dataset from various resources contain incomplete data. Next, normalization is employed to perform linear data transformation. It is also called min-max normalization where all attribute values are between the range of 0 and 1. Further, the pre-processed data are provided as input to the FE process.

3.3Textual feature extraction
The FND based text approaches employ original vector system. The approach is effectively employed for the identification of minimal and dissimilar sentences. Yet, the processing is complex for identifying the actual contents. The polysemy model is used for solving the above discussed problem and to identify the words before and after relations. To address this problem, this research applies HTBERT approach [42, 43] for FE, as represented by Equation 1.

$$h_i^t = BERT(t_i) \quad (1)$$

Where, h_i^t is the textual feature vector and t_i – ith is the input sentence.

The word2vec is a ML approach used for performing text embedding. Every word in text is depicted as a vector, permitting to estimate the similarity degree between the words and distance of two vectors. The aim of this approach is to combine maximum words but not merge these words. This technique involves two models: continuous bag-of-words (CBOW) for forecasting the target word wt, as formulated in Equation 2, and skip-gram which uses wt for the anticipation of framework, as illustrated in Equation 3.

$$L = \sum_{w_t \epsilon c} \log p(w_t | Context(w_t)) \quad (2)$$

$$L = \sum_{w_t \epsilon c} \log p(Context(w_t) | w_t) \quad (3)$$

Where, c represents the entire words in a training set. By utilizing the BERT model, the textual features are significantly extracted. After that, the parameters in BERT are tuned with the default values, alongside continuing the training process. According to the estimation results, an optimal hyperparameter is determined with the hyperparameter optimization model. This research utilizes grid search optimization algorithm for the determination of an optimal hyperparameter.

After that, the image feature is extracted by ResNet110 architecture and the detailed information is provided below.

3.4Image feature extraction

The FE technique is significantly implemented to extract crucial data features for the reduction of dimensionality. The aim of this approach is to reduce the large number of vital sources to explain numerous data. The greater amount of data supports the quality supplied for FE. The high-level image outcomes like shape, colour and structure are used for FE. The CNN architecture contains pooling and convolutional layer which have become most popular in computer vision, and are used for FE. For CNN, the feature map numbers are acquired by utilizing the convolutional task of the kernel as well as identifying the image perception features [44, 45]. The number of image features wholly concatenate the textual features illustrated through the feature vectors. This approach utilizes the outcome of CNN as low-level image feature set in this phase, which is then fused through altering the detection approach to distinguish the image's physical level determination.

This approach utilizes the ResNet110 architecture [46] as input for image encoding which is numerically represented in Equation 4 as:

$$h^v = ResNet110(v) \quad (4)$$

Where, v is the actual input image and h^v is an extracted visual semantic feature using ResNet110.

Further, semantic features are forwarded to the attention module to target the image regions. Therefore, an image feature distributes the weight for representing the model's importance when acquiring a visual model illustration, which is as expressed in Equations 5 to 7.

$$u^v = U^T \tanh(W^v h^v + b^v) \quad (5)$$

$$\alpha^v = \frac{\exp(u^v)}{\sum_i \exp(u^v)} \quad (6)$$

$$s^v = \sum_i \alpha_i h^v \quad (7)$$

Where, W^v is the weight matrix, b^v is a bias term, U^T is the transposed weight vector, and u is the scoring function which identifies an individual significance of the vector.

The word2vec [47, 48] is utilized for obtaining a better semantic data in image-to-sequence model, often used to develop visual characteristics for an image's features organization by text. This approach is typical for the utilization of embedding layer in text determination which forwards a language feature image to language level sequence, as formulated in Equation 8.

$$f^v = word2vec(s^v) \quad (8)$$

Where, s^v is the extracted language linguistic representation using ResNet110, and f^v is the visual language sequence by word2vec. The image features are significantly extracted using the ResNet110. Then, the visual tampering feature modules are extracted using the ELA algorithm and this is briefly explained in the following section.

3.5Visual tampering feature extraction module

In comparison to the image of a real news, a fake news image is repeatedly, maliciously spliced or has undergone a recompression many times because of its proliferation. The algorithm of ELA [49] focuses on malignant spliced as well as fake image attributes which is better than sending the image directly into the frequency domain by evaluation. The ELA is used to compress an image processing that is gradually changing. In a visual tampering FE module, the ELA transformation of an image is utilized. Later, ResNet110 extracts the image features as numerically shown in Equations 9 and 10.

$$v^{ela} = ELA(v) \quad (9)$$

$$h^{ela} = ResNet110(v^{ela}) \quad (10)$$

Where, v denotes the original input image, v^{ela} denotes an original image refined with ELA, and h^{ela} denotes the tampered feature by ResNet110. The visual tampering feature modules are efficiently extracted by the ELA algorithm. The extracted features of the text, image and visually tampered feature outcomes are then provided to the similarity measurement module.

3.6Similarity measurement module

In FND, false news is distinguished by measuring the cooperation between the data of text and images. The similarity module measurement is used to evaluate fake news of text and image likeness. The sub-networks of text and visual semantic features are certified by learning the basic representation patterns of image and text space. This method applies the fully connected (FC) layer to the sub-networks' final layer and treats sub-networks for transforming the last layer loads. This method efficiently generates the same depiction to a similar group of images as well as text samples. Later, cosine similarity is applied to estimate the similarity among the text and image [50], which is represented by Equations 11 and 12.

$$s = \frac{s^t . s^v}{\|s^t\| \times \|s^v\|} \qquad (11)$$

Where, s^t and s^v respectively denote a linguistic sequence of images and text.

$$p^s = sigmoid(s) \qquad (12)$$

Where, sigmoid is an activation method deployed to map the range from 0 to 1. A similarity measurement module measures the resemblance of an image and text data. Then, the outcome of this approach is provided to the multimodal approach for identifying the fake data in both text and image, further briefly explained in the following section.

3.7Multimodal approach

After acquiring the textual and visual feature depiction by using HTBERT and ResNet110, this research implements a feature fusion mechanism to attain a shared depiction to integrate an individual feature vector. This fusion captures the text and image data, providing a comprehensive set of features to detect fake news. The concatenated feature vectors serve as input to the classifier which also allow complementary data usage from both text and visual data for a commendable detection performance. Through leveraging the both types of features, FND systems captures the wide range of characteristics which denote the presence of fake news. The fusion mechanism enhances the relationship among textual and visual feature depictions. It is assumed that $(R_T) \in R_m$ for text feature vector, and $(R_V) \in R_n$ for visual feature vector. The general multimodal bilinear model is mathematically expressed in Equation 13.

$$R_{TV} = R_T^T W_i R_V \qquad (13)$$

Where, $W_i \in R^{mn}$ is the projection matrix, and R_{TV} is the bilinear model outcome. To minimize the parameter numbers, W_i is factorized as a low rank matrix, mathematically expressed in Equation 14.

$$R_{TV} = 1^T (U_i^T R_{T^o} V_i^T R_V) \qquad (14)$$

Where, U_i^T and V_i^T are the element-wise multiplication of two vectors, and they are redeveloped as two-dimensional matrices, $U' \in R^{m*ko}$ and $V' \in R^{n*ko}$, mathematically expressed in Equation 15.

$$R_{TV} = \frac{R_{TV}^T}{\|R_{TV}\|} \qquad (15)$$

In this approach, the feature fusion mechanism enhances the relationship between text and image features and provides suitable alignment.

The CNN-based multimodal approach utilizes the text and image's similar news for FND. Once the pre-processed data are forwarded to CNN, several tasks are imposed to the multimodal data. The CNN is performed in the unimodal structure which ignores the last multiple dense layers of rectified linear activation function (ReLU). CNN primarily labels the information provided to the convolutional layer and then forwards it to the hidden layer. It is eventually forwarded to the further input standard. A filter size (5 × 5) is used with stalk from 1 to 0 padding. A convolutional layer's outcome is forwarded to ReLU activation function and then, the max-pooling layer acquires a filter size (2 × 2), resulting in (278 × 278) dimensions. Then, the outcome of maxpooling is forwarded to the convolutional layer which is an input for the outcome channels of 6 and 3. These outcomes are then forwarded to the succeeding layer when the stride length, padding, as well as filter sizes are equal. The ReLU and max-pooling are required for subsequent feature maps from the convolutional layer. Therefore, the given input data acquires multiple feature maps (137 × 137). A textual data is set through CNN, rather than the dense outcome to the SoftMax layer; thus the output vector illustrating the text is associated with the vector obtained from CNN. Then, the vector is forwarded to dense layers among ReLU. Eventually, a logsoftamx layer is required and the possibilities logarithm is used for the provided input class examination. Therefore, fake news is significantly classified by using CNN. Finally, the output size of CNN is identical to the number of hidden units which is 256 with a batch size of 32 and the model is trained by using 10 epochs.

3.7.1Loss function (L)

The cross-entropy is an efficient probability distribution which compares the performance of prediction values. If the cross-entropy gets minimum accuracy and the prediction gets maximum accuracy, it becomes 0 and the prediction is efficient. In binary classification where the number of classes (M) equals 2, cross-entropy is estimated as given in Equation 16.

$$L = -(ylog(p) + (1-y)\log(1-p)) \qquad (16)$$

Where, y is the binary indicator (0 or 1), and p is the predicted probability observation o of class c.

3.8Dropout

Dropout is a regularization method for reducing overfitting problem in classifier which eliminates a complicated co-adaptation of training data. An idea of this method is to arbitrarily provide units from a network during in training. It is efficient in creating a model averaging with classifier. Every hidden neuron's FC layer outcome is advanced to zero with a 0.5 probability. As follows, the neurons are "dropped out", hence not contributing to backpropagation (BP). The FC layers are performed in training but, the results are multiplied through 0.5 outcomes.

Further, the evaluation and implementation of the suggested model is provided in the below section.

Algorithm: Grid Search Convolution Neural Network

Input: Normalized data

Output: Fake news detection

Begin

Step 1: Set input data weight: W_i, W_f, W_c, W_o

Step 2: Set Recurrent data Weight: R_i, R_f, R_c, R_o

Step 3: Set peephole weight: $V \epsilon R^N$

Step 4: Set Offset: $b_i, b_f, b_c, b_o \epsilon R^N$

Step 5: At time t, x_t is the input and y_t is the output of the node

Step 6: perform convolution operation at time t.

Step 7: ml.add (values of learning rates, optimization algorithm)

Step 8: gd-GdSearchCV (par_gd=par_gd, est=ml)

Step 9: gd_res=gd.fit(data samples of real and fake)

Step 10: Initialize the CNN (optimal CNN) with optimal parameters gd_res.

Step 11: $i_t = \sigma(W_i x_t + R_i h_{t-1} + b_i)$ is the output of the input data of optimal CNN at time t.

Step 12: $\acute{C}_{to}$, C_{to} is the input and cell structure of the optimal CNN node at time t, respectively, which are expressed as:

$$\acute{C} = tanh(W_c x_t + R_c h_{t-1} + b_c)$$

$$C = i_t \odot \acute{C}_t + f_t \odot c_{t-1}$$

Step 13: $O_t = (W_o x_t + R_o h_{t-1} + b_o)$ is the output of the CNN network

Step 14: The final output h_t of the node is expressed:

$$h = O_t \odot \tanh(C_t)$$

End

4.Experimental result

In this section, the performance and effectiveness of the proposed multimodal FND is evaluated with a fake news dataset named Fakeddit for both text and image of fake news.

4.1Dataset description

In this suggested method, the fake news dataset named Fakeddit [41] is used for training and testing the model, consisting of 1 million examples. The samples are based on 2, 3 and 6-way classification types by reserved supervision. In 2-way classification, the fake news articles are divided into two classes of fake and not fake. In 3-way classification, the fake news articles are divided into three classes of fake, satirical and not fake. Furthermore, in 6-way classification, the fake news articles are divided into 6 classes of satirical, not fake, misleading, fabricated, opinion and clickbait. This dataset includes a number of posts that are collected from the users of Reddit, containing a number of images, text, metadata and comments. The dataset is used to perform a fine-grained fake news classification, as well as to determine and identify if the news is real or fake. In this Fakeddit, every example has a label which distinguishes fake news in five categories. The dataset is classified into the partitions of training, testing and validation. This dataset contains two variety versions called unimodal which contain only text, while multimodal contains both text and images. This dataset contains 682,661 images of news and 2,90,000 text news. It encompasses a ratio of 2:3 for real and fake news. This denotes that the proposed approach is capable of simplifying efficiently the various levels of complexity and various types of fake news. Furthermore, the suggested method for detection of fake news through combining various features is significant and outstrips the existing others. The dataset has imbalance problem which also affects the classification task as it is difficult for those classes with fewer instances. Hence, the dataset is split into a ratio of 70:30 for both text and image. The 5,63,523 samples are utilized for training, 59,283 are used for validation, and 59,257 are utilized for testing.

4.2Evaluation metrics

Several performance metrics are utilized to validate the proposed method of multimodal FND. The employed hyperparameters have an initial learning rate of 0.001 and 0.01, the number of layers in the CNN are 256, feature vector size of 32, Adam optimizer with learning rate 0.01 is used to optimize the model, with momentum of 0.0 and 0.2 and decay rate of 0.001. The performance metrics accuracy, precision, recall, F1-score, mean absolute error (MAE) and root mean square error (RMSE) are estimated from the detection of all fake news. The mathematical expressions of these metrics are given in Equations 17 to 22:

- **Accuracy:** The ratio of all correct classifications to total number of classifications.

$$Accuracy = \frac{TP}{TP+FP} \times 100 \quad (17)$$

- **Precision:** The ratio of true positive over the classifications of all positive.

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (18)$$

- **Recall:** The proportion of original positives that are correctly classified.

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (19)$$

- **F1-score:** This combines precision and recall to give the average value of weight.

$$\text{F1-score} = \frac{2x\ Precision\ x\ Recall}{Precision + recall} \times 100 \quad (20)$$

- **MAE:** It is identified through a complete variation among predicted and actual values.

$$MAE = \frac{1}{K}\sum_{i=1}^{K}|\widehat{y_i} - y_i| \quad (21)$$

- **RMSE:** It is estimated through average error size and is disturbed by differences from an actual value.

$$RMSE = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(\widehat{y_i} - y_i)^2} \quad (22)$$

Where, TP is True Positive, FP is False Positive, TN is True Negative, and FN is False Negative.

4.3Quantitative and qualitative analysis

This section provides the performance analysis of an image CNN model in terms of achievable sum rate. The experimental model is performed using the CNN model for VGG16, VGG19, ResNet 50, and ResNet110.

Table 1 represents the performance analysis of various networks for FND of the image data. The existing methods such as VGG16, VGG19 and ResNet 50 are estimated and compared with the ResNet110 architecture. The ResNet110 for the image data enables much faster training at each layer and also achieves better accuracy results. The acquired outcomes prove that the ResNet110 attains commendable results based on the performance metrics: accuracy, precision, recall and F1-score with corresponding values of about 98.85%, 97.89%, 97.01%, and 98.32%, alongside MAE of 0.54 and RMSE of 0.57.

Table 1 Performance analysis of ResNet110 method with various methods network methods for image data

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MAE	RMSE
VGG16	92.91	91.65	91.68	91.75	0.84	0.91
VGG19	94.87	93.25	93.09	94.54	0.76	0.87
ResNet50	96.90	96.48	96.71	96.59	0.54	0.68
ResNet110	98.85	97.89	97.01	98.32	0.44	0.57

Table 2 exhibits the performance analysis of various networks for the FND of the text data. The existing methods which are MVAE, TextGCN, and RNN, are estimated and compared with the HTBERT. The HTBERT for the text data enables much faster training at each layer and also achieves better accuracy results. The acquired outcomes show that the HTBERT attains superior results on the basis of the performance metrics: accuracy, precision, recall and F1-score, with corresponding values of about 92.95%, 91.39%, 90.69% and 92.90%, alongside MAE of 0.53 and RMSE of 0.59. *Table 3* presents the performance evaluation of various networks for multimodal FND. The existing methods: NB+RF, NB+SVM and BiLSTM+CNN are analyzed and contrasted with the introduced HTBERT+ResNet110. The HTBERT+ResNet110 for the image and text data enables much faster training at each layer with commendable accuracy results. The acquired outcomes prove that the HTBERT+ResNet110 attains preferable results on the performance metrics of accuracy, precision, recall and F1-score with values of about 93.15%, 94.49%, 94.26% and 94.66%, alongside MAE of 0.51 and RMSE of 0.61. *Table 4* displays the outcomes of computation time as consumed by the proposed method, in contrast with the existing methods. It is hence derived that the HTBERT+ResNet110 method achieves a minimum

computational time of 300ms at the number of epochs being 150, which is preferable in contrast to the existing methods. In this research, the computational efficiency of the proposed method is analysed based on the computational time.

Table 2 Performance analysis of HTBERT method with various methods network methods for text data

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MAE	RMSE
MVAE	86.34	85.24	84.89	85.56	0.88	0.89
TextGCN	88.79	87.25	87.19	87.89	0.76	0.88
RNN	91.40	89.47	89.20	91.02	0.67	0.62
HTBERT	92.95	91.39	90.69	92.90	0.53	0.59

Table 1 Performance analysis of proposed method with various methods network methods for Multimodal data

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	MAE	RMSE
NB+RF	71.98	70.28	70.13	71.09	0.87	0.91
NB+SVM	74.06	73.25	82.19	73.89	0.72	0.83
BiLSTM+CNN	76.26	74.90	74.98	75.65	0.64	0.77
HTBERT+ResNet110	93.15	94.49	94.26	94.66	0.51	0.61

Table 4 Performance analysis of computational time

Methods	Computational time (ms)
NB+RF	600
NB+SVM	500
BiLSTM+CNN	400
HTBERT+ResNet110	300

4.4Comparative analysis

This section demonstrates the comparative analysis of the developed method against the existing methods. The comparative analysis is carried out with the utilization of the dataset, and the consideration of the accomplished results on the different performance metrics. *Table 5* demonstrates the comparative analysis of the proposed method alongside the previous models.

Table 5 Comparative Analysis of proposed method with existing methods for fake news

Author	Method	Dataset	Accuracy	Precision	Recall	F1-score
Liu et al. [21]	RCNN	Weibo	0.888	0.862	0.920	0.893
		Fakeddit	0.925	0.938	0.937	0.937
Ying et al. [23]	MMCN	Weibo	0.879	0.886	0.861	0.879
		Pheme	0.872	0.837	0.780	0.807
Wang et al. [25]	FMFN	Weibo	0.885	0.878	0.851	0.864
Proposed HTBERT+ResNet110	HTBERT+ResNet110	Fakeddit	0.931	0.944	0.942	0.946

5.Discussion

The obtained classification outcomes of the developed model are presented in *Tables 1* to *3*. The existing method, RCNN [21] deploys two datasets, Weibo and Fakeddit. On the Weibo dataset, it achieves an accuracy of 0.888 and precision of 0.862. Likewise, on the Fakeddit dataset, it achieves an accuracy of 0.925, precision of 0.9444, recall of 0.942 and F1-score of 0.946. The MMCN [23] method makes use of the Weibo and Pheme fake news dataset, where the obtained results on the Weibo dataset is 0.879 for accuracy, 0.886 for precision, 0.861 for recall and 0.879 for F1-score. Whereas, on the Pheme dataset, the obtained results are: 0.872 of accuracy, 0.837 of precision, 0.780 of recall and 0.807 of F1-score. Furthermore, the FMFN [25] method utilizes the Weibo dataset. The obtained results on the Weibo dataset are: 0.885 for accuracy, 0.878 for precision, 0.851 for recall and 0.864 for F1-score. The number of posts on various events is imbalanced on the Fakeddit dataset. Due to this issue, the learned text features majorly focus on particular events only. It was complex for earlier textual modality approaches to extract the features of transferability among various events. Therefore, the textual performance is seen to be minimum for every approach. As a result, the HTBERT is proposed as a powerful tool for FE in text. The pre-trained HTBERT architecture follows the transformer model, in which the multi-head attention is deployed for preserving the semantic relations between words. The developed multimodal HTBERT+ResNet110 deploys the Fakeddit fake news dataset and achieves better results on various performance metrics given as:

0.931 for accuracy, 0.944 for precision, 0.942 for recall and 0.946 for F1-score. The proposed HTBERT+ResNet110 achieves better accuracy results of 0.931 when compared to the existing methods such as RCNN, MMCN and FMFN. However, RCNN attains poor accuracy result of 0.925 as compared to the HTBERT+resNet110. The proposed method is proven to be most efficient in overcoming the event imbalance impact through the employment of the event domain adaptation network that estimates variation among various events and ignores the individual features of every event, to build the model's performance. Nevertheless, developing an effective hand-crafted feature needs great knowledge of the related areas and of events. In the meantime, this method depends on hand-crafted features, where the acquired feature vector's robustness is not sufficient as it has no knowledge of FND.

5.1Limitations

The proposed HTBERT+ResNet110 method is introduced for FND in social media. This method is improved by integrating the name entity identification techniques which find the significant features present in the fake news by differentiating them from the original data. The proposed HTBERT+ResNet110 method extracts only the significant knowledge from the image and text fake news. The limitation identified for this method is that it does not perform on various real-time applications, but only considers the FND. This limitation of the proposed method can be overcome in future works.

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

6.Conclusion and future work

The FND in social media becomes challenging, because of which various tools are developed for the detection of fake news. In this work, the multimodal FND is proposed for determining the correctness of fake news. This method utilizes the multimodal approach on both text and image news for identifying if the news is real or fake. This method employs two classification methods: HTBERT classifier for text classification and ResNet110 for image classification. This developed model makes use of the Fakeddit dataset for an effective multimodal detection of fake news. As opposed to the previous methods, the proposed multimodal-based CNN combines text and image plans to detect the fake news. The proposed model exhibits a classification accuracy of 0.931, precision of 0.944, recall of 0.942, F1-score of 0.946. In the future, the developed method will be encompassed to optimize a method for feature fusion approach in numerous applications.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

Data availability

The datasets generated during and/or analyzed during the current study are available in [Fakeddit] dataset repository: https://paperswithcodecom/dataset/fakeddit.

Author's contribution statement

Chetan Agrawal: Conceptualization, investigation, data collection, interpretation of result, writing – original draft. **Anjana Pandey:** Conceptualization, investigation, data collection, writing – original draft. **Sachin Goyal:** Conceptualization, writing-review, and supervision.

References

[1] Katariya P, Gupta V, Arora R, Kumar A, Dhingra S, Xin Q, et al. A deep neural network-based approach for fake news detection in regional language. International Journal of Web Information Systems. 2022; 18(5/6):286-309.

[2] Fu L, Peng H, Liu S. KG-MFEND: an efficient knowledge graph-based model for multi-domain fake news detection. The Journal of Supercomputing. 2023; 79(16):18417-44.

[3] Dixit DK, Bhagat A, Dangi D. An accurate fake news detection approach based on a levy flight honey badger optimized convolutional neural network model. Concurrency and Computation: Practice and Experience. 2023; 35(1):e7382.

[4] Katarya R, Dahiya D, Checker S. Fake news detection system using featured-based optimized msvm classification. IEEE Access. 2022; 10:113184-99.

[5] Lai CM, Chen MH, Kristiani E, Verma VK, Yang CT. Fake news classification based on content level features. Applied Sciences. 2022; 12(3):1-21.

[6] Liao Q, Chai H, Han H, Zhang X, Wang X, Xia W, et al. An integrated multi-task model for fake news detection. IEEE Transactions on Knowledge and Data Engineering. 2021; 34(11):5154-65.

[7] Meel P, Vishwakarma DK. HAN, image captioning, and forensics ensemble multimodal fake news detection. Information Sciences. 2021; 567:23-41.

[8] Singhal S, Pandey T, Mrig S, Shah RR, Kumaraguru P. Leveraging intra and inter modality relationship for multimodal fake news detection. In companion proceedings of the web conference 2022 (pp. 726-34). ACM.

[9] Papadopoulos SI, Koutlis C, Papadopoulos S, Petrantonakis PC. VERITE: a robust benchmark for multimodal misinformation detection accounting for

unimodal bias. International Journal of Multimedia Information Retrieval. 2024; 13(1):1-15.
[10] Raza S, Ding C. Fake news detection based on news content and social contexts: a transformer-based approach. International Journal of Data Science and Analytics. 2022; 13(4):335-62.
[11] Cui X, Yang L. Fake news detection in social media based on multi-modal multi-task learning. International Journal of Advanced Computer Science and Applications. 2022; 13(7):1-7.
[12] Dong D, Lin F, Li G, Liu B. Similarity-aware attention network for multimodal fake news detection. Third international conference on big data & artificial intelligence & software engineering 2022 (pp.76-85).
[13] Ying Q, Hu X, Zhou Y, Qian Z, Zeng D, Ge S. Bootstrapping multi-view representations for fake news detection. In proceedings of the AAAI conference on artificial intelligence 2023 (pp. 5384-92).
[14] Hakak S, Alazab M, Khan S, Gadekallu TR, Maddikunta PK, Khan WZ. An ensemble machine learning approach through effective feature extraction to classify fake news. Future Generation Computer Systems. 2021; 117:47-58.
[15] Meel P, Vishwakarma DK. A temporal ensembling based semi-supervised ConvNet for the detection of fake news articles. Expert Systems with Applications. 2021; 177:115002.
[16] Mughaid A, Al-zu'bi S, Al AA, Al-amrat R, Alajmi R, Zitar RA, et al. An intelligent cybersecurity system for detecting fake news in social media websites. Soft Computing. 2022; 26(12):5577-91.
[17] Zhou Y, Yang Y, Ying Q, Qian Z, Zhang X. Multi-modal fake news detection on social media via multi-grained information fusion. In proceedings of the international conference on multimedia retrieval 2023 (pp. 343-52). ACM.
[18] Raj C, Meel P. ConvNet frameworks for multi-modal fake news detection. Applied Intelligence. 2021; 51(11):8132-48.
[19] Kaliyar RK, Goswami A, Narang P. EchoFakeD: improving fake news detection in social media with an efficient deep neural network. Neural Computing and Applications. 2021; 33:8597-613.
[20] Song C, Ning N, Zhang Y, Wu B. A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. Information Processing & Management. 2021; 58(1):102437.
[21] Liu P, Qian W, Xu D, Ren B, Cao J. Multi-modal fake news detection via bridging the gap between Modals. Entropy. 2023; 25(4):1-17.
[22] Segura-bedmar I, Alonso-bartolome S. Multimodal fake news detection. Information. 2022; 13(6):1-16.
[23] Ying L, Yu H, Wang J, Ji Y, Qian S. Multi-level multi-modal cross-attention network for fake news detection. IEEE Access. 2021; 9:132363-73.
[24] Guo Y, Song W. A temporal-and-spatial flow based multimodal fake news detection by pooling and attention blocks. IEEE Access. 2022; 10:131498-508.
[25] Wang J, Mao H, Li H. FMFN: fine-grained multimodal fusion networks for fake news detection. Applied Sciences. 2022; 12(3):1-12.
[26] Jing J, Wu H, Sun J, Fang X, Zhang H. Multimodal fake news detection via progressive fusion networks. Information Processing & Management. 2023; 60(1):103120.
[27] Kumari R, Ekbal A. Amfb: attention based multimodal factorized bilinear pooling for multimodal fake news detection. Expert Systems with Applications. 2021; 184:115412.
[28] Rai N, Kumar D, Kaushik N, Raj C, Ali A. Fake news classification using transformer based enhanced LSTM and BERT. International Journal of Cognitive Computing in Engineering. 2022; 3:98-105.
[29] Li S, Yao T, Li S, Yan L. Semantic-enhanced multimodal fusion network for fake news detection. International Journal of Intelligent Systems. 2022; 37(12):12235-51.
[30] Mehta D, Dwivedi A, Patra A, Anand KM. A transformer-based architecture for fake news classification. Social Network Analysis and Mining. 2021; 11:1-2.
[31] Palani B, Elango S, Viswanathan KV. CB-fake: a multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT. Multimedia Tools and Applications. 2022; 81(4):5587-620.
[32] Kaliyar RK, Goswami A, Narang P. FakeBERT: fake news detection in social media with a BERT-based deep learning approach. Multimedia Tools and Applications. 2021; 80(8):11765-88.
[33] Aslam N, Ullah KI, Alotaibi FS, Aldaej LA, Aldubaikil AK. Fake detect: a deep learning ensemble model for fake news detection. Complexity. 2021; 2021:1-8.
[34] Choudhary A, Arora A. Linguistic feature based learning model for fake news detection and classification. Expert Systems with Applications. 2021; 169:114171.
[35] Xue J, Wang Y, Tian Y, Li Y, Shi L, Wei L. Detecting fake news by exploring the consistency of multimodal data. Information Processing & Management. 2021; 58(5):102610.
[36] Guo Y. A mutual attention based multimodal fusion for fake news detection on social network. Applied Intelligence. 2023; 53(12):15311-20.
[37] Yang P, Ma J, Liu Y, Liu M. Multi-modal transformer for fake news detection. Mathematical Biosciences and Engineering: MBE. 2023; 20(8):14699-717.
[38] Sastrawan IK, Bayupati IP, Arsa DM. Detection of fake news using deep learning CNN–RNN based methods. ICT Express. 2022; 8(3):396-408.
[39] Al OA, Khotanlou H, Mansoorizadeh M, Zabihzadeh D. Multimodal fake-news recognition using ensemble of deep learners. Entropy. 2022; 24(9):1-16.
[40] Uppada SK, Patel P. An image and text-based multimodal model for detecting fake news in OSN's. Journal of Intelligent Information Systems. 2023; 61(2):367-93.

[41] https://paperswithcodecom/dataset/fakeddit.. Accessed 23 April 2024.
[42] Keya AJ, Wadud MA, Mridha MF, Alatiyyah M, Hamid MA. AugFake-BERT: handling imbalance through augmentation of fake news using BERT to enhance the performance of fake news classification. Applied Sciences. 2022; 12(17):1-21.
[43] Shishah W. Fake news detection using BERT model with joint learning. Arabian Journal for Science and Engineering. 2021; 46(9):9115-27.
[44] Liu Y, Pu H, Sun DW. Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices. Trends in Food Science & Technology. 2021; 113:193-204.
[45] Barbhuiya AA, Karsh RK, Jain R. CNN based feature extraction and classification for sign language. Multimedia Tools and Applications. 2021; 80(2):3051-69.
[46] Guo N, Gu K, Qiao J, Bi J. Improved deep CNNs based on nonlinear hybrid attention module for image classification. Neural Networks. 2021; 140:158-66.
[47] Zhao W, Zhu L, Wang M, Zhang X, Zhang J. WTL-CNN: a news text classification method of convolutional neural network based on weighted word embedding. Connection Science. 2022; 34(1):2291-312.
[48] Zhang F. A hybrid structured deep neural network with Word2Vec for construction accident causes classification. International Journal of Construction Management. 2022; 22(6):1120-40.
[49] Zhang Y, Shi T, Lu ZM. Image splicing detection scheme based on error level analysis and local binary pattern. Journal of Network Intelligence. 2021; 6:303-12.
[50] Singh R, Singh S. Text similarity measures in news articles by vector space model using NLP. Journal of The Institution of Engineers (India): Series B. 2021; 102:329-38.

Chetan Agrawal is pursuing PHD in CSE at University Institute of Technology Rajiv Gandhi Proudyogiki Vishwavidyalaya (UIT - RGPV), Bhopal, He Studied Master of Engineering in CSE at TRUBA Institute of Engineering & Information Technology Bhopal. He has studied his Bachelor of Engineering in CSE at BANSAL Institute of Science & Technology Bhopal. Currently, He is working as Assistant professor in the CSE department at RADHARAMAN Institute of Technology & Science Bhopal M.P. India. His research area of interest is Social Network Analysis, Data Analytics, Machine Learning, Cyber Security, Network Security, Wireless Networks, and Data Mining.
Email: chetan.agrawal12@gmail.com

Anjana Pandey received her MCA degree from GGDU University in Bilaspur, and her M.Tech in Computer Science from Banasthali Vidyapith in Rajasthan. She earned her Ph.D. in Computer Application in 2011 from MANIT, Bhopal, Madhya Pradesh, India. Currently, she is working as an Associate Professor in the Department of Information Technology at the University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (Madhya Pradesh), India. She has published over 30 research papers in various international and national journals and conferences, including SCIE and Scopus-indexed journals. She is currently guiding six students in their Ph.D. programs.
Email: anjanapandey@rgpv.ac.in

Sachin Goyal received his Bachelor's degree in Computer Science and Engineering from ITM Gwalior, his M.Tech. in Information Technology from Samrat Ashok Technological Institute, Vidisha, Bhopal, and his Ph.D. in Information Technology in 2013 from Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, Madhya Pradesh, India. He is currently an Associate Professor in the Department of Information Technology at the University Institute of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal (Madhya Pradesh), India. He has published over 30 research papers in various international and national journals and conferences, including SCIE and Scopus journals. He is guiding six Ph.D. students and serves as a reviewer for Applied Soft Computing, Elsevier, IET Taylor and Francis, and many other international journals and conferences. Previously, he taught at the National Law Institute University, Bhopal, under the M.S. program in Cyber Law and Information Security.
Email: sachingoyal@rgpv.ac.in

Appendix I

S. No.	Abbreviation	Description
1	ABM-CNN-RNN	Attention based Multi-level CNN-Recurrent Neural Network
2	AMFB	Attention-based Multimodal Factorized Bilinear Pooling
3	BP	Backpropagation
4	BERT	Bidirectional Encoder Representations from Transformers
5	Bi-LSTM	Bi-Directional LSTM
6	CB	CapsNet and BERT
7	CBOW	Continuous Bag-of-Words
8	CNN	Convolutional Neural Network
9	DL	Deep Learning
10	ELA	Error Level Analysis
11	FE	Feature Extraction
12	FMFN	Fine-Grained Multimodal Fusion Network
13	FND	Fake News Detection
14	FC	Fully Connected
15	GRU	Gated Recurrent Unit

16	HTBERT	Hyperparameter Tuning based Bidirectional Encoder Representations from Transformers
17	ISOT	Indian Society of Organ Transplantation
18	LSTM	Long Short-Term Memory
19	MAE	Mean Absolute Error
20	MANN	Mutual Attention Neural Network
21	MCNN	Multimodal Consistency Neural Network
22	ML	Machine Learning
23	MPFN	Multimodal Progressive Fusion Network
24	NLP	Natural Language Processing
25	RCNN	Recurrent Convolutional Neural Network
26	ReLU	Rectified Linear Activation Function
27	RMSE	Root Mean Square Error
28	TGA	Thermogravimetric Analysis