IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 26 September 2023, accepted 3 December 2023, date of publication 12 December 2023,
date of current version 20 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3341870

RESEARCH ARTICLE

A Multi-Kernel Optimized Convolutional Neural Network With Urdu Word Embedding to Detect Fake News

KHURRAM ZAHEER, MUHAMMAD RAMZAN TALIB, MUHAMMAD KASHIF HANIF, AND MUHAMMAD UMER SARWAR
Department of Computer Science, Government College University Faisalabad, Faisalabad 38000, Pakistan
Corresponding author: Muhammad Ramzan Talib (ramzan.talib@gcuf.edu.pk)

ABSTRACT One of the biggest threats to international trade, journalism, and democracy is fake news on social media, which also causes substantial collateral damage. The manipulation of digital news to spread misinformation has become a common practice for personal benefits or relief. Therefore, developing an automated system that can detect fake news before publication is crucial. This study proposes a three-level methodology with a new model called Multi-Kernel Optimized Convolutional Neural Network (MOCNN) to investigate its effectiveness for fake news detection. The parameters of the proposed model have been optimized using the grid search technique. We evaluated ten different deep learning models on two benchmark datasets of fake news articles and compared their performance with the proposed model. Finding the best model with good accuracy performance is the primary goal of this paper. F-measure and accuracy are used to evaluate and compare the classification performance of these deep learning models. Our proposed model achieves 85.8% and 68.2% accuracy and 85.8% and 67.7% F1-measure on UFN and BET, respectively. Experimental results confirm that the proposed model performs better than other models on UFN and BET datasets.

INDEX TERMS Fake news, Urdu fake news, machine learning, ensemble learning, dagging.

I. INTRODUCTION

Information purposely prepared and distributed to convey a misleading impression about a person, place, or thing is considered deceptive. A piece of news that is blatantly and demonstrably false is regarded as fake news [1]. One of the most damaging types of deception is fake news, in which the source of the story is unclear, and someone has changed the content of the official press [2]. As the Internet and web-based technologies have developed, people increasingly rely on social networking sites (like YouTube, Facebook, and Twitter) and electronic news blogs (digital newspapers and magazines) to access information. Real digital news can be manipulated easily and quickly and disseminated to spread false information among communities to gain a few benefits or find relief.

The associate editor coordinating the review of this manuscript and approving it for publication was Xianzhi Wang.

As a result of the quick distribution of fake news through these sources, the accuracy of the content on these news blogs and social networking sites is in doubt. Even though posting and propagating false information is wrong, people still do so for various reasons. To increase sales of their products, businesses exploit incorrect information. In the presidential election of the USA in 2016, politicians and their followers disseminated false information to affect the election results [3]. Therefore, the need for reliable, efficient, and effective ways to automatically detect fake news is increasing. Since 2016, researchers, companies, and societies have started paying more and more attention to fake news detection from online news platforms.

With over a hundred million native speakers, Urdu ranks as the world's ninth most widely spoken language [4]. Compared to other languages, such as English, Urdu has a relatively limited range of text-processing resources like datasets, stemmers, lemmatizers, and word-embedding. There are two main reasons to design an automated fake news detection system for Urdu. First, every language has its features. Urdu has different features like left-to-right writing direction, no capitalization, cursive text, etc., making Urdu text unique to English and other languages. Second, the preprocessing techniques are different for each language, like stopwords, stemming, tokenizer [5]. Although English is an international language, local languages like Urdu cannot be ignored in designing systems that detect fake news effectively. The most effective automated fake news detection system should be capable of detecting fake news in various languages [2]. It is also crucial to design an automatic system that can quickly identify fake news posted by a user for languages with limited resources like Urdu, remove it, and show an alert or ban the user from online news applications.

A significant research issue is creating automatic, trustworthy, and precise methods for detecting fraudulent information on social media. Since 2017 Facebook, Twitter, and Google have begun to focus on developing systems to identify and counter fake news. Major systems have previously used machine learning techniques to assist users in identifying fake news in German, Latin, Slavic [6], Portuguese [2], Urdu [7], and other languages [8], [9]. Machine learning-based techniques have a number of shortcomings:1) perform well at a small subset of features, 2) rely on manually created features that require domain-specific knowledge, 3) choose effective features from high dimensional features, and 4) performance difference across models when applied to the same corpus due to variations in the size and distribution of the corpus' instances. In several empirical experiments, deep learning models have outperformed more conventional machine learning techniques [10], [11].

Deep learning is a branch of machine learning that uses the artificial neural network structure with many layers like an input, one or more hidden, and an output. CNN and RNN are two popular deep-learning architectures [12]. CNN extracts feature by employing convolution layers, pooling layers, and fully linked layers. They are considered good at learning non-sequential patterns using 1-D kernels in convolutional layers, which can capture local context information among neighboring words [13]. However, CNN do not capture long-term dependencies among words since they process the text independently [11]. On the other hand, RNN and its variations, such as LSTM and BiLSTM, are well-suited for learning time-series data and can record inter-word dependencies for extended periods by using memory cells to keep the network's state [14].

Shallow CNN extracts same-length features because of the kernels of the same length in the convolution layer. In text processing, past studies show that multiple kernels of the same size perform poor than the variable size kernels in the convolutional layer [10], [15]. This study proposes a three-level methodology to detect fake news using deep learning models. It includes dataset design and preprocessing, model training and validation, and testing the model. We propose a new Multi-kernel Optimized Convolutional Neural Network (MOCNN) model that uses multiple variable-length kernels in the convolutional layer. After text preprocessing (tokenizing, cleaning, stopwords, and removing rare words), MOCNN takes the news article as input from the Urdu word embedding layer that outputs it to the convolutional layer, where a word is represented as a vector. Multiple variable-length kernels in the convolutional layer allow the model to extract variable-length semantic features that better detect fake news articles [16], [17]. The output of the convolutional layer is the multiple feature maps (local features) of variable size (one from each kernel) that are merged together to form a final feature map (global features) after the max-pooling operation. This paper uses ten models based on CNN and RNN architectures as baseline models to process Urdu text, analyze, and compare their performance with the proposed model. The results show that the proposed model has the best performance on both datasets compared to other models.

The significant technical contributions of this paper are summarized as follows:

- We propose a three-level methodology to detect fake news from Urdu articles.
- To detect fake news, we propose the MOCNN model, which uses multiple kernels of variable length to extract high-level and low-level features from the input data.
- We optimize the proposed MOCNN model by applying the grid search technique to pick the optimal values of the model parameters.
- Furthermore, we used two benchmark datasets for performance comparison. Accuracy and the F-measure were used to evaluate the performance of the MOCNN model.
- We evaluate the proposed model against the performance of ten deep-learning models, such as CNNs, RNNs, and Transformers. The proposed model outperforms the others on both datasets.

The remainder of the paper is structured as follows: In Section II, we provide a literature analysis on detecting fake news for Urdu and other resource-poor languages. Three-level methodology proposed in this study is described level by level in Section III, along with the baseline models used for comparison with the proposed mode. The architecture of the proposed deep learning model is discussed in detail in Section IV. Section V details the procedure followed to fine-tune the proposed model's parameters. Section VI summarizes and analyzes the experimental results from the ten baseline and proposed models. Section VII brings the study to a conclusion and future work.

II. RELATED WORK

The Internet and web-based applications such as news blogs, Twitter, Facebook, and instant messaging offer much knowledge on many topics. When compared to other sources, the dissemination of fake news utilizing these applications is both quicker and more convenient [18]. Due to the large amount of labor, time, and resources required, early detection and blocking of false news stories before they are uploaded online is impossible to accomplish manually. Therefore, creating methods for accurately and effectively using machine learning and deep learning techniques to detect fake news is crucial.

Classifiers that utilize machine learning have demonstrated outstanding performance in a variety of text classification tasks [19], including document classification [20], email classification [21], sentiment analysis [22], the detection of fraudulent reviews [23] and abusive comment detection [24]. In many research for resource-rich languages like English, a range of machine learning techniques were employed to automatically categorize fake news and real news extracted from a news dataset. [25], [26]. The linguistic resources for the English language, such as data crawling and preprocessing, are readily available to the public [27], however, this is not the case for languages that have a scarcity of linguistic resources.

Machine learning, deep learning, and ensemble learning are the primary focuses of the strategies used to detect fake news on social media [28] employed an ensemble learning technique gradient boosting algorithm to detect fake news from a multiclass dataset and shows that gradient boosting outperforms the machine learning models. [6] investigated four machine learning models KNN, SVM, NB, and RF for fake news detection from Latin, Germanic, and Slavic languages. Results on four datasets shows that no model outperforms the others on all the datasets. Same results were reported by [29] on the Urdu text classification. Similarly, [30] presents a method to detect fake news that is language-independent. They employed English, Spanish, and Portuguese datasets and four machine-learning models.

The average detection accuracy was 85% using random forest. Reference [31] proposed a multi-view attention network to detect fake news from social media and compared its performance with seven deep learning models to show the effectiveness of the proposed model. Similarly, [32] propose EchoFakeD a deep learning model that achieved 92% accuracy on news dataset. Another method, Multi-level word features based CNN, was proposed to achieve high classification performance than the base-line models [33].

The ability to identify fake news in Urdu languages with limited resources is not good as it is in the English-speaking world. Machine learning methods for languages with limited resources have been the subject of a few studies in the past. Lack of linguistic resources, such as annotated news corpus, is a significant barrier [5]. Contrary to English, there is no publicly accessible corpus with millions of documents. Numerous text-processing research are restricted to datasets with small sizes and a small number of news articles. It is expensive, time-consuming, and requires specialized knowledge to manually create an annotated corpus for a language with limited resources [7]. Because there are not many datasets available, researchers have been employing augmented datasets that have been primarily machine translated from English into other languages [1], [34].

TABLE 1. Datasets used for fake news identification for languages with limited resources.

Corpus Name	Language	Real	Fake	References
--	Chinese	131	187	[35]
DECOUR	Italian	1202	945	[36]
--	English and Spanish	100	100	[37]
CSI	Dutch	270	270	[38]
Fake.Br	Portuguese	3600	3600	[2]
Bend the Truth	Urdu	500	400	[7]
MT	Urdu	400	400	[1]
UFN	Urdu	1032	968	[34]
Covid19Fake	Arabic	3567	1387	[40]
BET	Urdu	500	400	[5]
RU-FRDC	Roman Urdu	3714	1326	[4]
	Thai	13816	13816	[8]

Table 1 compares the dataset in detail used for fake news detection in languages other than English. On a small corpus of 318 news stories in Chinese language text, [35] performed SVM-based classification. Italian news stories were divided into 'legitimate' and 'fake' categories using a corpus of 2147 articles [36]. Syntactic, lexical, and semantic variables from a small corpus of 200 French news articles were employed by [37] to identify fake news. In order to train and evaluate the model to predict false news articles from a Dutch corpus of 540 news items, [38] utilized an SVM classifier with ten-fold cross-validation. All of these datasets, except Portuguese, are noted to be smaller, with only a few hundred news articles. Additionally, these datasets are too small to form a conclusive or general judgment regarding the effectiveness of any machine-learning classification method because the data determines how well machine-learning algorithms work. (dimensionality, size, and distribution of examples in the corpus) [39].

Although the deep learning and machine learning approaches and their respective stated results were reviewed above, such approaches mainly concentrate on news articles written in popular languages like English, Spanish, etc. The methods for detecting fake news written in Urdu are minimal, and there is a clear need for additional study and analysis to achieve a higher level of performance. This study aims to investigate the effectiveness of various well-known deep learning models and propose an effective deep learning model to detect fake news in Urdu.

III. METHODOLOG

In this section, we discuss our proposed three-level methodology to investigate the effectiveness of the deep learning models in detecting fake content from Urdu fake news articles. Figure 1 shows the three levels in graphical form. The three-level method starts with input dataset preprocessing, training the models on the training dataset, and testing these

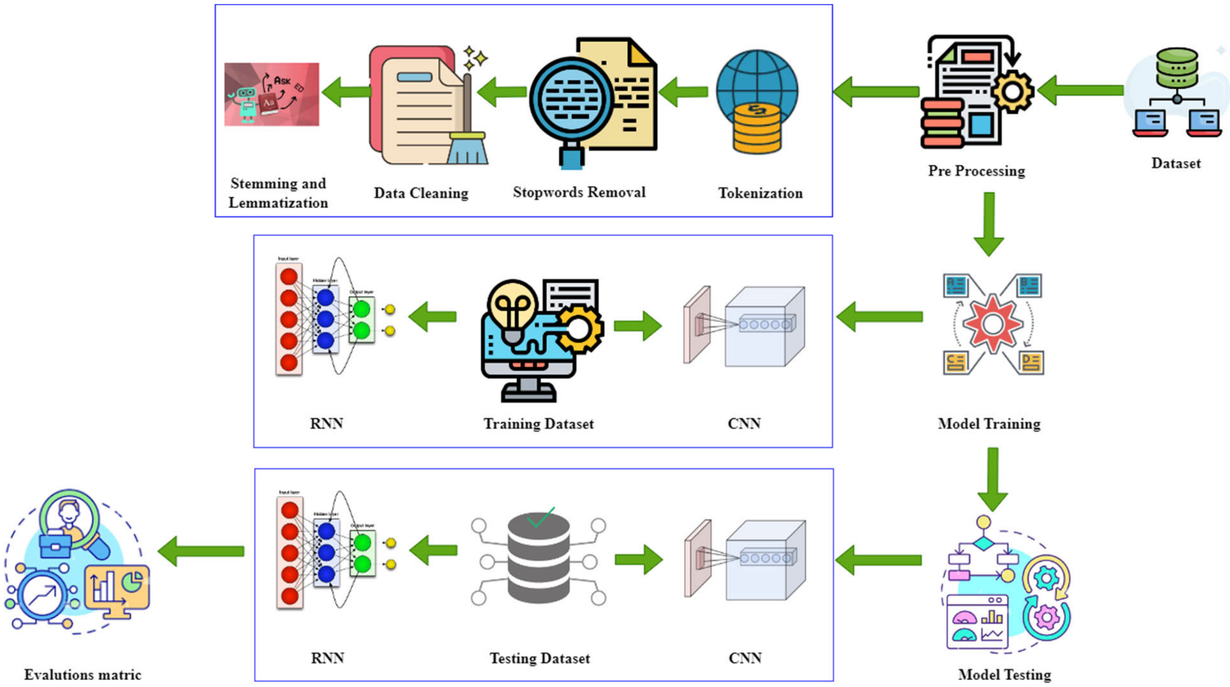


FIGURE 1. Procedure of preprocessing steps, training, and testing the deep learning models.

trained models on the tested dataset. We concisely discuss each level and the methods used at the level below.

A. LEVEL-0: FAKE NEWS DATASETS AND LANGUAGE PREPROCESSING

Since Urdu has few linguistic resources, the need for annotated datasets is the primary barrier to conducting brief research for fake news detection. Two small datasets from recent studies are publicly available to detect fake news. Both Urdu Fake News (UFN) [34] and Bend the Truth (BET) each have 2,000 news pieces and 900 news articles [7]. Table 2 provides a statistical comparison of the UFN and BET datasets. For the aim of detecting fake news, both datasets are challenging for deep learning models. UFN is more significant than BET because of its large vocabulary size. BET is more difficult because of its small vocabulary and short-length articles. Because of the small size, datasets are still imbalanced, although the difference between the number of fake and real articles in both datasets is insignificant. It is essential to use these datasets in this study because small and imbalanced datasets are more challenging deep learning models because insufficient training lead to model overfitting.

Deep learning models cannot process the news article directly. Every piece in the dataset must be preprocessed and converted into a form easily readable by the model. In this study, the preprocessing steps performed on the UFN and BET datasets are the same as those performed in [5] for fake news detection and are shown in Figure 2. First, we tokenized the text of an article using special characters (like space, tab, dot, and comma). Second, we clean the tokenized piece by deleting special characters, non-Urdu words and characters, and URLs. Finally, we eliminate the most frequent (stopwords) and infrequent words (rare words) to decrease the dataset size. Researchers found that the deep learning model's effectiveness was enhanced by removing both stopwords and rare words [41]. A list of Urdu stopwords from GitHub[1] is utilized to eliminate stopwords. Three hundred thirty-nine stopwords are included in the original list. Because many study papers claim that stemming decreases the classification performance, such as Urdu [42], Roman Urdu [43], Arabic

TABLE 2. Statistics of the UFN and BET dataset.

Properties	Urdu Fake News (UFN)		Bend the Truth (BET)	
	Real	Fake	Real	Fake
Total articles	1,032	968	500	400
Maximum Length Article	6,068	7,045	1,153	2,159
Minimum Length Article	25	25	59	57
Vocabulary Size	954,254	1,147,547	120,394	184,023
Average Article Length	961	1230	372	300

[1]https//github.com/urduhack/urdu-stopwords

FIGURE 2. Procedure of text preprocessing including tokenizing text, cleaning text, and removing stopwords and rare words.

[44], and Turkish [45], the stemming operation has not been carried out on both datasets.

Figure 3 presents a word cloud comparing the content of fake news to that of real news. To construct this word cloud, we first eliminated all the special characters, stopwords, and rare words. Because of this, both data sets contain actual terms that describe the articles in real or fake class documents. In the BET dataset, پاکستان, بھارت, چین,امریکہ are the prominent words of different countries in fake news. Political parties and their voters spread propaganda against the government and its policies. In real news, the words are different that the fake news like پاکستان, مطابق, سال, ٹیم. The word cloud also illustrates that the news articles in both fake and actual classes are taken from political, sports, business, and health domains. In UFN dataset, most important words are ٹرمپ, امریکہ, کلنٹن, ریپبلکن, صدر, and مہم. These words depict that the articles in both classes belong to the domain of politics. This dataset includes the news articles collected during the US presidential election. In presidential election of the USA in 2016, politicians and their followers disseminate false information to affect the election results [3].

B. LEVEL-1: TRAINING DEEP LEARNING MODELS

After preprocessing, a dataset is often split into training, validation, and testing subsets to train, validate, and test the deep learning models. A 60:20:20 split was employed to divide the dataset into training, validation, and testing sets for this analysis. The embedding layer receives the input text and output to hidden layers in the convolutional layer. To the best of our knowledge, no standard pre-trained word embedding like a glove is publicly available. Therefore, the word embedding used in the proposed architecture is trained on the given Urdu news article dataset. The size of the embedding layer is taken as a hyperparameter. Convolutional layers in CNN and memory units in RNN extract valuable features. We trained CNN, DPCNN, LSTM, BiLSTM, Transformer, and five other deep learning models on two datasets. The hyper-parameters like batch size, number of epochs, and dropout of these deep learning models have been fine-tuned on both datasets before the final experiments.

C. LEVEL-2: TESTING THE TRAINED MODELS

After completing the training of deep learning models, the testing dataset is used to test the performance of these models. We used two well-known performance measures to test our proposed and other deep learning models: F-Measure and Accuracy. Experimental results on Urdu [41] and English [46] text classification shows that accuracy is not a suitable performance metric when the examples in the dataset are not equally distributed. F-measure is a good performance measure when we have a dataset with imbalanced class distribution because it gives equal value to each class [47]. These performance measures are well-known measures in text classification using deep learning. F-measure is also known as the harmonic mean of precision and recall. The following equation can be used to compute it.

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{1}$$

The Recall is also referred to as true positive rate and sensitivity. Recall measures the completeness of a classification system. It is the ratio of the true positive (TP) news articles to the sum of the true positive and false negative (FN) news articles. The equation to calculate Recall is given below:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

FN is the sum of all real news articles in the dataset that the classifier incorrectly labeled as fake. TP indicates the total number of fake news articles that the model successfully predicts as fake news articles. Precision measures the exactness of the classifier. It is the TP to the TP plus false positive (FP) ratio. FP is the total number of fake news in the corpus but is predicted as real news by the classifier. The following equation is used to find Precision.

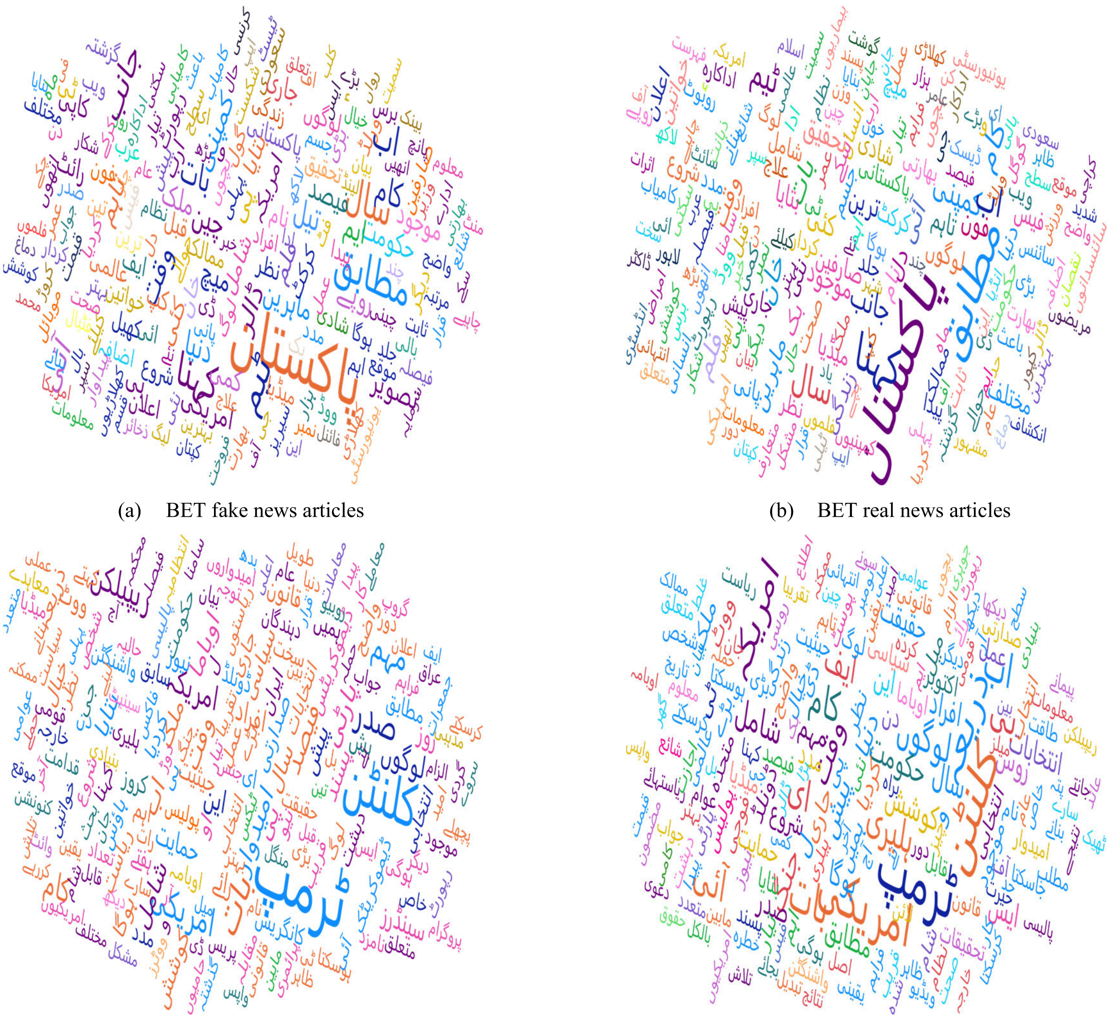
$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Accuracy is the standard by which classification performance is measured. It indicates the ratio of correctly labeled news articles to the dataset's total number of news articles. It can be calculated as given below:

$$Accuracy = \frac{No.\ of\ correctly\ classified\ articles}{Total\ number\ of\ articles} \tag{4}$$

IV. PROPOSED MODEL

The convolutional neural network has effectively performed both image and text processing. Our proposed multi-kernel optimized convolutional neural network (MOCNN) is a CNN-based model. The proposed model has a single channel



(a) BET fake news articles

(b) BET real news articles

(c) UFN fake news articles

(d) UFN real news article

FIGURE 3. Word cloud of the BET and UFN datasets. Word clouds of (a) and (b) shows the important words in BET fake news and real news. (c) and (d) shows the important words in fake and real news articles of UFN.

to take the input to the word embedding layer. We could not locate any publicly available, industry-standard, pre-trained word embedding comparable to the glove. Therefore, the proposed architecture of the CNN model employs a word embedding that was trained using Urdu, as mentioned earlier in news article datasets. The size of the word embedding in the embedding layer is taken as a hyper-parameter and chosen using a grid search method. In the convolutional layer, many feature maps are generated, one for each convolutional kernel size employed. The 1-max pooling algorithm is performed on each feature map. After that, a feature vector for the penultimate layer is created by concatenating all of the feature maps in the previous layers. The softmax layer serves as the final layer, and its purpose is to classify the news article being analyzed into one of two categories. Figure 4 presents the framework that served as the basis for this study. The proposed model employs a 1-D word order structure in the convolutional layer to learn an effective text feature representation from the training dataset. It does this by employing convolutional kernels of varying sizes (window size), which are then used to extract variable-length features or n-grams from the text. It captures the local elements or relationship between neighboring words regarding context windows. Then it uses pooling layers to extract global features from the data. As a result, it has the potential to achieve better results than a model that employs numerous kernels of the same size in the convolutional layer [15]. When employing pooling layers, it is also helpful to discover short-range and long-range relations inside both the short and the long text [17].

Formally, we can define the architecture of the proposed model. Suppose that x is an input news article that contains L words and that d is the length of the word vector then $x \in R^{L \times d}$ is the input and $x_i \in R^d$ is a word in x where i represents the word position and d is the dimension of the word vector. A news article of length n (padded by a special

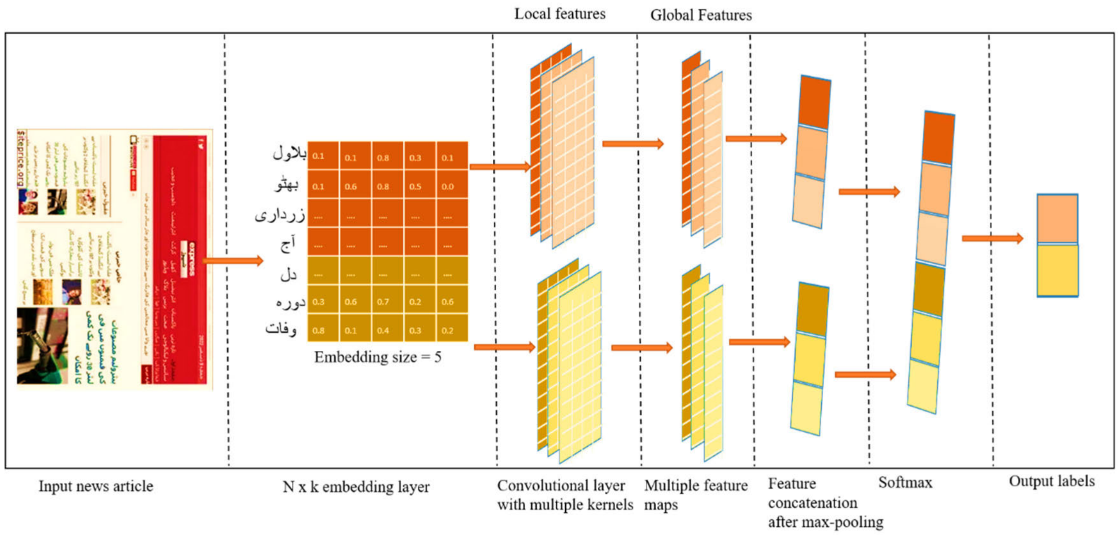


FIGURE 4. Architecture of the proposed model where embedding layer convert an Urdu news article to word embedding. Multiple feature maps are obtained from multiple kernels of different size and concatenated to form a final feature map. Final label of the article is assigned by the softmax layer.

token if necessary) is then represented as the concatenation of the word vectors. A news article x can be represented as below where $+$ shows the concatenation operation.

$$x_{1:n} = x_1 + x_2 + \ldots + x_n \tag{5}$$

Within the convolutional layer, a kernel f with a length k (the total number of words) is selected., where vector $n \in R^{k\times d}$ represents a kernel for the convolution operation. Each time, the kernel will analyze the words beginning at position j and ending at position (j+k-1) sequentially. The equation below represents window w_i.

$$w_i = [x_i + x_{i+j} + \ldots + x_{j+k-1}] \tag{6}$$
$$c_i = f(w.x_{i:i+h-1} + b) \tag{7}$$

In this case, f could be a sigmoid or other non-linear function, The term $b \in R$ is biased, and the operation . is element-wise multiplication. All the features taken from all the available word-windows are compiled into a feature map c, which is then calculated using the equation below.

$$c = [c_1, c_2, \ldots, c_{n-h+1}] \tag{8}$$

where $c \in R^{n-h+1}$. After the convolutional operation, the pooling operation is performed by the pooling layer. In this study, To prevent overfitting from noisy text and choose the top k (most relevant) features from the feature map for subsequent processing, we employed k max-pooling. Maxpooling operation can be represented as equation given below.

$$c' = max\{c_1, c_2, \ldots, c_{n-h+1}\} \tag{9}$$

During the process of pooling, both the order in which words appear and the information between words are considered. A fully connected softmax layer serves as CNN's output layer. This layer provides a probability distribution value based on the article labels. Softmax operation is as given below.

$$y_j = w_j y_{j-1} + b_j \tag{10}$$

where y_j is a vector representing the output of the softmax layer, y_{j-1} is a vector representing the output of the pooling layer, w_j is a transition matrix representing the softmax layer, and b_j is a bias factor. The following equation illustrates how the probability distribution is distributed across all of the article labels:

$$P(i|t; \emptyset) = \frac{exp(y_{ij})}{\sum_{k=1}^{n} exp(y_{ij})} \tag{11}$$

Network overfitting is possible because to the extensive hyper-parameter tuning. In order to solve the issues of overfitting and a large number of hidden units and the connection between them, dropout regularization is applied to a fully connected layer.

V. NETWORK MODEL'S PARAMETERS OPTIMIZATION

For good performance, the hyper-parameters of a classifier must be optimized. The goal of parameter optimization is to maximize the performance of a classifier by determining the optimal values for its parameters for a given dataset. In this study, we used a grid search approach to fine-tune the hyper-parameters of our model. These parameters are initially set to the identical values used in [10]. Initialized values to hyper-parameters are listed in Table 3. We padded smaller articles with the word 'UNK' up to the dataset's maximum length to guarantee that all pieces have a fixed length and produce batches of a fixed size. By mapping each article word to an integer index in the vocabulary, we could also establish a language containing all unique observations, including "UNK." We could express each article using this method as a fixed-length vector of integers.

TABLE 3. Parameters and their initial values used to optimize the MOCNN model.

Parameter	Value	Parameter	Value
Batch Size	50	Kernel Size	3,4,5
Embedding Size	300	No. of kernels	100
No. of Epoch	50	Dropout	0.5
Activation Function	ReLu	Training-Testing	80-20

To represent words in the embedding layer, they are assigned a vector with real-numbered values, and this vector's size can influence the model's performance. A fixed-size, pre-defined word embedding is not utilized, and the size of the embedding vector is determined as a hyper-parameter. Results in Figure 5 (a) show that the embedding size is independent of the dataset size. Embedding size 32 achieves maximum accuracy on both datasets.

The convolution operation utilizes several kernels to compute feature maps. The number of kernels and kernel size can impact the model's performance. Large kernels can cause slow training, while small kernels may miss crucial distinguishing information, leading to decreased performance [48]. Further, finding kernels with an appropriate size is a time-consuming and challenging task. Figure 5 (b) compares five different sizes of kernels. The comparison shows that kernel size 1, 2 outperforms the others on BET datasets, while kernels 1, 3 outperform the others on the UFN dataset. Multiple kernels can exploit features of different n-grams. In other words, applying filters of 1 and 3 lengths in the convolutional layer will result in two feature maps of unigrams and trigrams. Multisized kernels perform better than the same size kernels, as concluded in [10] and [33]. Figure 5 (c) shows the number of kernels analyzed during experiments, and the best value is 128 from both datasets. Batch size refers to the total number of news articles given as input to the network in a single iteration processing. A large batch size makes training longer and consumes more GPU memory. A batch size 32 performed the best when we tested our model with five various batch sizes (16, 32, 48, 64, and 80). This analysis is given in Figure 5 (d).

When the training data is small, dropout regularization is an efficient approach to stop overfitting and noise. When the dataset is complicated and noisy, it is necessary to either expand the size of the dataset or reduce the number of hidden units used to compute features. Dropout prohibits inactive units from participating in the calculation for the following iteration by deactivating or eliminating them from the hidden layer. Figure 5 (e) shows that the model performs best for both datasets with 0.5 dropouts.

An epoch is a model passing through every news article in a corpus or batch in both a forward and backward pass. More epoch are needed for CNN to learn a large, complicated, and noisy dataset. Still, caution should be used because doing so too often might cause the network to become overfit, which causes it to perform well on the training dataset but poorly on the testing dataset. Figure 5 (f) demonstrates that our model performs best on both datasets with 30 epochs.

The pooling layer reduces the size of the feature map by selecting the most significant feature from neighboring features. This layer enables higher-level layers to choose more abstract or global values. In our study, we employed a 1-max-pooling operation in the pooling layer. The Adagrad algorithm optimizes the error and adjusts the learning rate based on the parameters. Gradient descent is used to optimize network training. The softmax layer is used as a probabilistic function to calculate a confidence score for the classification decision. After extensive testing on two datasets, the improved hyper-parameter values are shown in Table 4.

TABLE 4. Optimized parameters of MOCNN on both datasets.

Parameters	UFN	BET
Embedding Size	32	32
Kernel Size	3,4,5	3,4,5
No. of kernels	128	128
No. of Epochs	30	30
Activation Function	ReLu	ReLu
Dropout	0.5	0.5
Batch Size	32	32

VI. RESULTS AND DISCUSSIONS

This section provides the experimental results in graphical and tabular forms and a brief discussion. In this section, we compare the performance of deep learning models to classify fake news articles. After finding the optimal parameters, we performed the final experiments by applying our proposed model to both datasets. We have also fine-tuned the parameters of other deep-learning models. Experimental results on the UFN corpus show that the proposed model MOCNN outperforms the others in detecting fake news. Obtained accuracy values are shown in Figure 6. Because it uses the multiple kernels of size three to exploit the features of three lengths (or three words), and it captures the contextual relationship among the words. Therefore, it performs better than the shallow CNN and the DPCNN. Our model achieves the best accuracy score of 85.2%. LSTM model with attention layer performs better than other models except for MOCNN. The transformer model gives the worst performance among the deep learning models.

Accuracy is not a good performance measure when the dataset is imbalanced. Therefore, we also evaluate the performance of our models to detect fake news using F-measure. Figure 7 shows that, with accuracy, the proposed model shows significant performance among the other models by

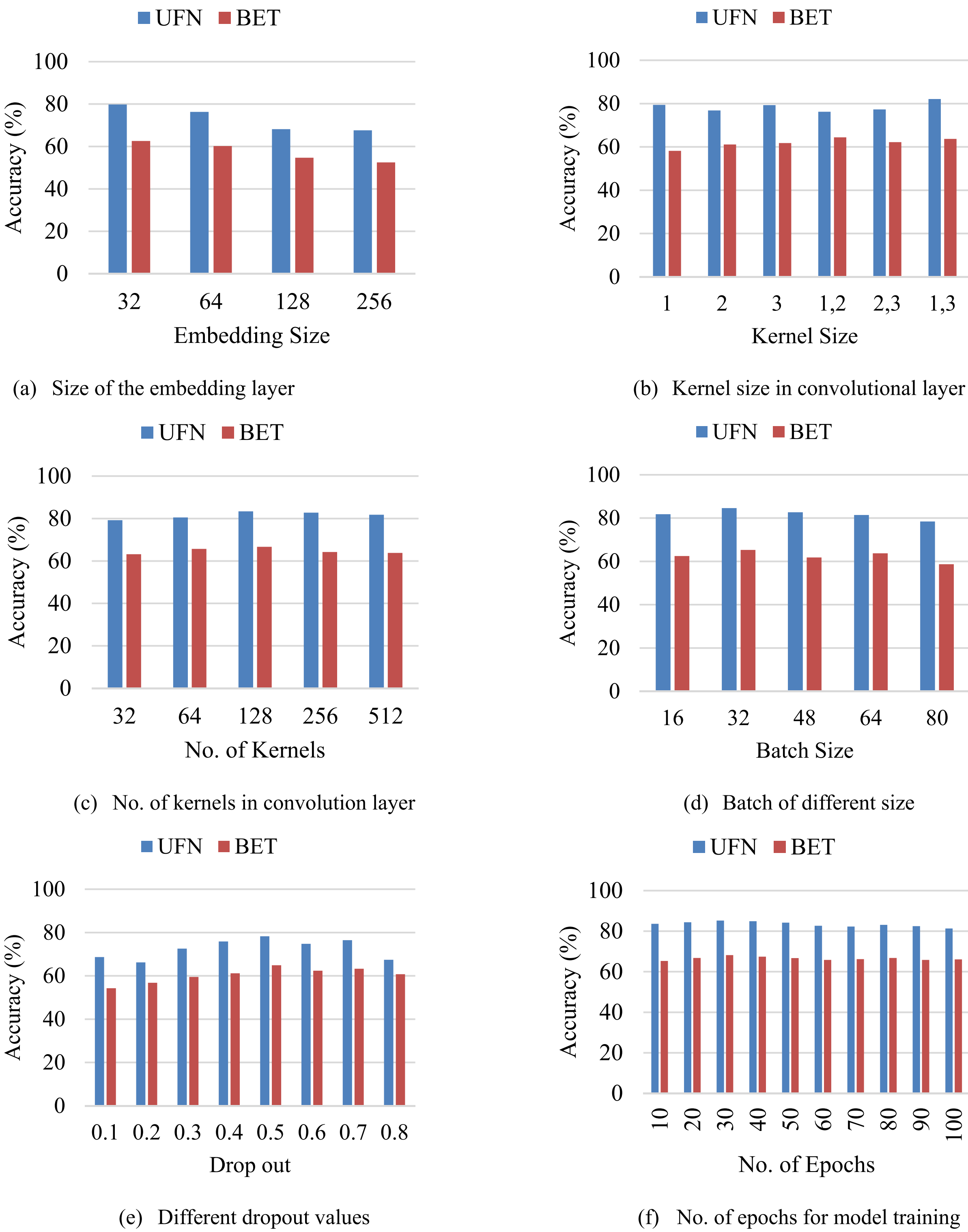


FIGURE 5. Analysis of the different values of hyper-parameters of the proposed model to choose the beast values for high classification performance.

achieving an 85.8% f-measure score of 0.5% higher than the accuracy score. The second prominent f-measure score is 83.2%, acquired by the LSTM with an attention mechanism. LSTM can capture more rich and semantic information from the text, and the attention mechanism retains the essential features by setting different weights.

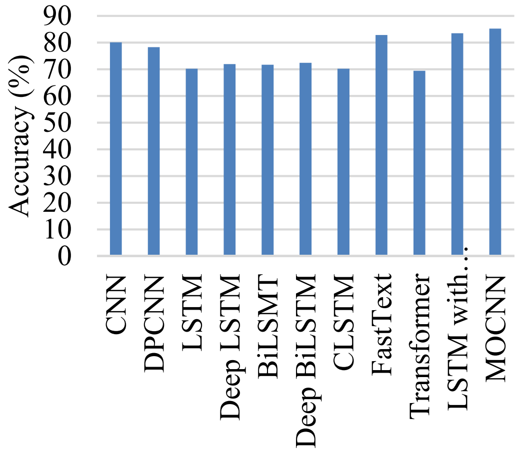


FIGURE 6. Accuracy achieved by the MOCNN and other deep learning models on UFN.

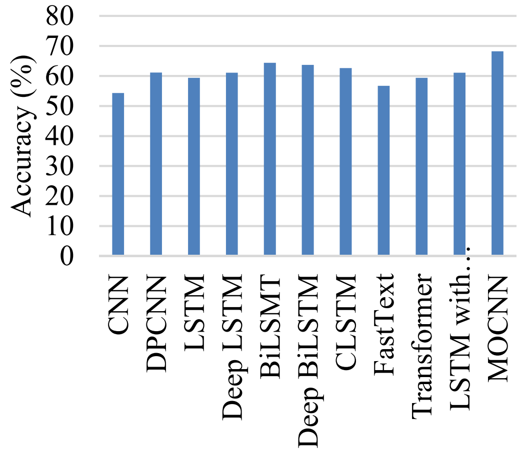


FIGURE 8. Comparison of the proposed model with other deep learning models on BET dataset using accuracy scores.

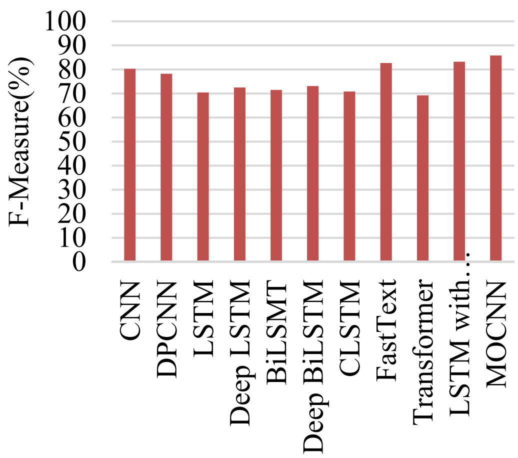


FIGURE 7. Comparison of the proposed model with other deep learning models on UFN dataset using F-measure scores.

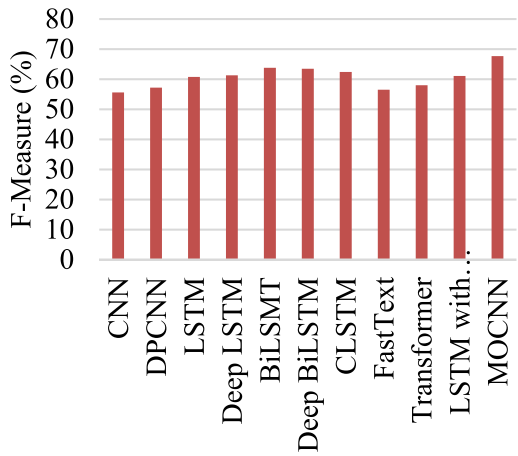


FIGURE 9. Comparison of the proposed model with other deep learning models on BET dataset using F-measure scores.

The proposed model MOCNN outperforms the others on the BET corpus and achieves a 68.8% accuracy score. Results given in Figure 8 show that the proposed model's architecture is good at detecting fake news from small and large datasets. LSTM and BiLSTM and their variants models obtained higher accuracy scores than CNN and its models. But this is different on UFN, a larger dataset than BET, where LSTM and its variant models show the worst performance. Even the attention mechanism within LSTM and Transformer could not be more helpful for small datasets.

F-measure scores of all the models are given in Figure 9 for comparison on the BET dataset. Here, the results are similar to those shown in Figure 4 and discussed above. Again, the LSTM and its variants are more outstanding than the CNN and its variants. But, the proposed model outperforms the other models and achieved 67.7% F-measure scores.

On both datasets, the results of all the models have been summarized in Table 5. Lower results of deep learning models are noted on the BET than UFN because of its small size and small vocabulary. It shows that deep learning models perform well if the corpus size is large. Although we optimized the hyperparameters of our models on both datasets, the insufficient number of training examples caused to decrease the performance. Shallow CNN and Transformer models have obtained the minimum accuracy and showed poor performance on BET and UFN, respectively.

When we compare the variants models of RNN and CNN architectures, we notice that CNN and its variants outperform the large-size UFN dataset. In contrast, RNN and its variants show dominant performance on small-size BET datasets. Even the models with deep layered architecture like DPCNN, Deep LSTM, and Deep BiLSTM have not performed better on both datasets. The attention mechanism with LSTM shows significant performance among the other nine models, but the same is not valid for the Transformer model with a multi-head attention mechanism. The transformer delivers the worst performance on both datasets. When we analyze the accuracy and F-measure scores in UFN and BET columns, there is a difference of a small margin among both performance measures. Usually, this happens when we use a balanced

TABLE 5. Summary of the results achieved by the deep learning models.

Classifier	UFN		BET	
	F-measure	Accuracy	F-measure	Accuracy
CNN	80.3	80.1	55.6	54.3
DPCNN	78.2	78.3	57.2	61.11
LSTM	70.4	70.2	60.8	59.4
Deep LSTM	72.5	71.9	61.3	61.1
BiLSMT	71.5	71.7	63.8	64.4
Deep BiLSTM	73.1	72.4	63.5	63.7
CLSTM	70.8	70.2	62.4	62.6
FastText	82.7	82.8	56.5	56.7
Transformer	69.2	69.4	58.0	59.4
LSTM with Attention	83.2	83.5	61.1	61.1
MOCNN	85.8	85.2	**67.7**	**68.2**

dataset. In this study, both datasets are imbalanced, but their imbalanced level is small because of their small size. This margin is small for a small BET dataset and is large on large UFN dataset.

VII. CONCLUSION

In this study, we propose a three-level methodology and a CNN-based optimized model's architecture for fake news detection from the news articles in Urdu language text. We invite the researchers to focus on the resource-poor language Urdu by providing an effective deep learning model and methodology and by comparing the results in depth of the proposed model with ten other deep learning models using two performance evaluation methods. We systematically define a way to determine the optimized parameters for deep learning models. The experiments were performed diligently to achieve the objectives of this study. We conclude that the proposed model outperforms the other ten deep learning models on both datasets. On UFN, it reaches 85.2% accuracy and 85.8% F-measure. On BET, it achieves 67.7% F-measure and 68.2% accuracy. RNN-based models like LSTM and BiLSTM and their variants significantly perform on small BET datasets.

On the other hand, CNN-based models like shallow CNN and DPCNN are significantly better than RNN-based models on large UFN datasets. Attention mechanism with LSTM proven better than the multi-head attention in Transformer. Models with deep architectures like DPCNN, DeepLSTM, and DeepBiLSTM achieve high accuracy and f-measure scores than the simple deep learning models.

It would be essential to future work to design a large dataset including Urdu text and images and apply the proposed model with two-channel input to detect fake news. We also plan to use the proposed model to detect fake content from short text like tweets or reviews.

REFERENCES

[1] M. Amjad, G. Sidorov, and A. Zhila, "Data augmentation using machine translation for fake news detection in the Urdu language," in *Proc. 12th Lang. Resour. Eval. Conf.*, May 2020, pp. 2530–2535. [Online]. Available: https://www.aclweb.org/anthology/2020.lrec-1.308

[2] R. M. Silva, R. L. S. Santos, T. A. Almeida, and T. A. S. Pardo, "Towards automatically filtering fake news in Portuguese," *Expert Syst. Appl.*, vol. 146, May 2020, Art. no. 113199, doi: 10.1016/j.eswa.2020.113199.

[3] Y.-F. Huang and P.-H. Chen, "Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms," *Expert Syst. Appl.*, vol. 159, Nov. 2020, Art. no. 113584, doi: 10.1016/j.eswa.2020.113584.

[4] U. Hayat, A. Saeed, M. H. K. Vardag, M. F. Ullah, and N. Iqbal, "Roman Urdu fake reviews detection using stacked LSTM architecture," *Social Netw. Comput. Sci.*, vol. 3, no. 6, pp. 1–9, Sep. 2022, doi: 10.1007/s42979-022-01385-6.

[5] A. Rafique, F. Rustam, M. Narra, A. Mehmood, E. Lee, and I. Ashraf, "Comparative analysis of machine learning methods to detect fake news in an Urdu language corpus," *PeerJ Comput. Sci.*, vol. 8, Jun. 2022, Art. no. e1004, doi: 10.7717/peerj-cs.1004.

[6] P. H. A. Faustini and T. F. Covões, "Fake news detection in multiple platforms and languages," *Expert Syst. Appl.*, vol. 158, Nov. 2020, Art. no. 113503, doi: 10.1016/j.eswa.2020.113503.

[7] M. Amjad, G. Sidorov, A. Zhila, H. Gómez-Adorno, I. Voronkov, and A. Gelbukh, "'Bend the truth': Benchmark dataset for fake news detection in Urdu language and its evaluation," *J. Intell. Fuzzy Syst.*, vol. 39, no. 2, pp. 2457–2469, Aug. 2020, doi: 10.3233/jifs-179905.

[8] P. Meesad, "Thai fake news detection based on information retrieval, natural language processing and machine learning," *Social Netw. Comput. Sci.*, vol. 2, no. 6, pp. 1–7, Nov. 2021, doi: 10.1007/s42979-021-00775-6.

[9] M. Alkhair, K. Meftouh, K. Smaili, and N. Othman, "An Arabic corpus of fake news: Collection, analysis and classification," in *Proc. ICALP*, 2019, pp. 292–302, doi: 10.1007/978-3-030-32959-4_21.

[10] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, and M. Fayyaz, "Exploring deep learning approaches for Urdu text classification in product manufacturing," *Enterprise Inf. Syst.*, vol. 16, no. 2, pp. 223–248, Feb. 2022, doi: 10.1080/17517575.2020.1755455.

[11] M. N. Shah and A. Ganatra, "A systematic literature review and existing challenges toward fake news detection models," *Social Netw. Anal. Mining*, vol. 12, no. 1, p. 168, Dec. 2022, doi: 10.1007/s13278-022-00995-5.

[12] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid CNN-RNN based deep learning approach," *Int. J. Inf. Manage. Data Insights*, vol. 1, no. 1, Apr. 2021, Art. no. 100007, doi: 10.1016/j.jjimei.2020.100007.

[13] M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar, and M. S. Rahman, "A comprehensive review on fake news detection with deep learning," *IEEE Access*, vol. 9, pp. 156151–156170, 2021, doi: 10.1109/ACCESS.2021.3129329.

[14] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of CNN and RNN for natural language processing," 2017, *arXiv:1702.01923*.

[15] C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, "A C-LSTM neural network for text classification," 2015, *arXiv:1511.08630*.

[16] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. AbdelMajeed, and T. Zia, "Abusive language detection from social media comments using conventional machine learning and deep learning approaches," *Multimedia Syst.*, vol. 28, no. 6, pp. 1925–1940, Apr. 2021, doi: 10.1007/s00530-021-00784-8.

[17] H. Saleh, A. Alharbi, and S. H. Alsamhi, "OPCNN-FAKE: Optimized convolutional neural network for fake news detection," *IEEE Access*, vol. 9, pp. 129471–129489, 2021, doi: 10.1109/ACCESS.2021.3112806.

[18] R. K. Kaliyar, A. Goswami, P. Narang, and V. Chamola, "Understanding the use and abuse of social media: Generalized fake news detection with a multichannel deep neural network," *IEEE Trans. Computat. Social Syst.*, pp. 1–10, Nov. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9956917, doi: 10.1109/TCSS.2022.3221811.

[19] S. Lakhotia and X. Bresson, "An experimental comparison of text classification techniques," in *Proc. Int. Conf. Cyberworlds (CW)*, Oct. 2018, pp. 58–65, doi: 10.1109/CW.2018.00022.

[20] T. Zia, M. Akhter, and Q. Abbas, "Comparative study of feature selection approaches for Urdu text categorization," *Malays. J. Comput. Sci.*, vol. 28, pp. 93–109, Jan. 2015.

[21] A. Akhtar, G. R. Tahir, and K. Shakeel, "A mechanism to detect Urdu spam emails," in *Proc. IEEE 8th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Oct. 2017, pp. 168–172, doi: 10.1109/UEMCON.2017.8249019.

[22] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowl. Inf. Syst.*, vol. 60, no. 2, pp. 617–663, Aug. 2019, doi: 10.1007/s10115-018-1236-4.

[23] R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers," *Inf. Process. Manage.*, vol. 56, no. 4, pp. 1234–1244, Jul. 2019, doi: 10.1016/j.ipm.2019.03.002.

[24] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, and M. T. Sadiq, "Automatic detection of offensive language for Urdu and Roman Urdu," *IEEE Access*, vol. 8, pp. 91213–91226, 2020, doi: 10.1109/ACCESS.2020.2994950.

[25] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Inf. Sci.*, vol. 497, pp. 38–55, Sep. 2019, doi: 10.1016/j.ins.2019.05.035.

[26] G. Gravanis, A. Vakali, K. Diamantaras, and P. Karadais, "Behind the cues: A benchmarking study for fake news detection," *Expert Syst. Appl.*, vol. 128, pp. 201–213, Aug. 2019, doi: 10.1016/j.eswa.2019.03.036.

[27] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Inf. Process. Manage.*, vol. 57, no. 2, Mar. 2020, Art. no. 102025, doi: 10.1016/j.ipm.2019.03.004.

[28] R. K. Kaliyar, A. Goswami, and P. Narang, "Multiclass fake news detection using ensemble machine learning," in *Proc. IEEE 9th Int. Conf. Adv. Comput. (IACC)*, Dec. 2019, pp. 103–107, doi: 10.1109/IACC48062.2019.8971579.

[29] T. Zia, Q. Abbas, and M. P. Akhtar, "Evaluation of feature selection approaches for Urdu text categorization," *Int. J. Intell. Syst. Appl.*, vol. 7, no. 6, pp. 33–40, May 2015.

[30] H. Q. Abonizio, J. I. de Morais, G. M. Tavares, and S. Barbon Junior, "Language-independent fake news detection: English, portuguese, and Spanish mutual features," *Future Internet*, vol. 12, no. 5, p. 87, May 2020, doi: 10.3390/FI12050087.

[31] S. Ni, J. Li, and H.-Y. Kao, "MVAN: Multi-view attention networks for fake news detection on social media," *IEEE Access*, vol. 9, pp. 106907–106917, 2021, doi: 10.1109/ACCESS.2021.3100245.

[32] R. K. Kaliyar, A. Goswami, and P. Narang, "EchoFakeD: Improving fake news detection in social media with an efficient deep neural network," *Neural Comput. Appl.*, vol. 33, no. 14, pp. 8597–8613, Jul. 2021, doi: 10.1007/s00521-020-05611-1.

[33] Q. Hu, Q. Li, Y. Lu, Y. Yang, and J. Cheng, "Multi-level word features based on CNN for fake news detection in cultural communication," *Pers. Ubiquitous Comput.*, vol. 24, no. 2, pp. 259–272, Apr. 2020, doi: 10.1007/s00779-019-01289-y.

[34] M. P. Akhter, J. Zheng, F. Afzal, H. Lin, S. Riaz, and A. Mehmood, "Supervised ensemble learning methods towards automatically filtering Urdu fake news within social media," *PeerJ Comput. Sci.*, vol. 7, p. e425, Mar. 2021, doi: 10.7717/peerj-cs.425.

[35] H. Zhang, S.-D. Wei, H.-Y. Tan, and J.-H. Zheng, "Deception detection based on SVM for Chinese text in CMC," in *Proc. 6th Int. Conf. Inf. Technol., New Generat.*, Apr. 2009, pp. 481–486, doi: 10.1109/ITNG.2009.66.

[36] T. Fornaciari and M. Poesio, "Automatic deception detection in Italian court cases," *Artif. Intell. Law*, vol. 21, no. 3, pp. 303–340, Sep. 2013, doi: 10.1007/s10506-013-9140-4.

[37] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," in *Proc. COLING*, 2018, pp. 3391–3401. [Online]. Available: https://aclanthology.org/C18-1287/

[38] B. Verhoeven and W. Daelemans, "CLiPS stylometry investigation (CSI) corpus: A Dutch corpus for the detection of age, gender, personality, sentiment and deception in text," in *Proc. 9th Int. Conf. Lang. Resour. Eval., LREC 2014*, vol. 2014, pp. 3081–3085.

[39] K. Pham, D. Kim, S. Park, and H. Choi, "Ensemble learning-based classification models for slope stability analysis," *Catena*, vol. 196, Jan. 2021, Art. no. 104886, doi: 10.1016/j.catena.2020.104886.

[40] W. Shishah, "JointBert for detecting Arabic fake news," *IEEE Access*, vol. 10, pp. 71951–71960, 2022, doi: 10.1109/ACCESS.2022.3185083.

[41] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, A. Mehmood, and M. T. Sadiq, "Document-level text classification using single-layer multisize filters convolutional neural network," *IEEE Access*, vol. 8, pp. 42689–42707, 2020, doi: 10.1109/ACCESS.2020.2976744.

[42] A. Ali and M. Ijaz, "Urdu text classification," in *Proc. 7th Int. Conf. Frontiers Inf. Technol.*, 2009, doi: 10.1145/1838002.1838025.

[43] K. Mehmood, D. Essam, and K. Shafi, "Sentiment analysis system for Roman Urdu BT–intelligent computing," in *Advances in Intelligent Systems and Computing*, vol. 2019, pp. 29–42.

[44] A. Ayedh, G. Tan, K. Alwesabi, and H. Rajeh, "The effect of preprocessing on Arabic document categorization," *Algorithms*, vol. 9, no. 2, p. 27, Apr. 2016, doi: 10.3390/a9020027.

[45] M. Çağataylı and E. Çelebi, "The effect of stemming and stop-word-removal on automatic text classification in Turkish language," in *Neural Information Processing* (Lecture Notes of Computer Science of Springer), 2015, pp. 168–176. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-26532-2_19

[46] H. Chen, S. McKeever, and S. J. Delany, "A comparison of classical versus deep learning techniques for abusive content detection on social media sites," in *Social Informatics*, 2018, pp. 117–133. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-01129-1_8

[47] K. Stapor, P. Ksieniewicz, S. García, and M. Woźniak, "How to design the fair experimental classifier evaluation," *Appl. Soft Comput.*, vol. 104, Jun. 2021, Art. no. 107219, doi: 10.1016/j.asoc.2021.107219.

[48] M. H. Goldani, R. Safabakhsh, and S. Momtazi, "Convolutional neural network with margin loss for fake news detection," *Inf. Process. Manage.*, vol. 58, no. 1, Jan. 2021, Art. no. 102418, doi: 10.1016/j.ipm.2020.102418.

KHURRAM ZAHEER received the M.S. degree in software engineering from Government College University Faisalabad, in 2017, where he is currently pursuing the Ph.D. degree in computer science. He has been with the School Education Department for the last 15 years. He has published two several publications in reputed journals. His research interests include natural language processing, social media analysis, text classification, data mining, and machine learning.

MUHAMMAD RAMZAN TALIB received the Ph.D. degree from the University of Bayreuth, Germany. He is currently a Professor with the Department of Computer Science, Government College University Faisalabad, Pakistan. He has been the Head of the Department of Computer Science. He has experience teaching for more than 20 years. He has published many research articles in high-quality journals. His research interests include databases and information systems, data mining, data warehousing, and text classification.

MUHAMMAD KASHIF HANIF received the Ph.D. degree from the Hamburg University of Technology, Germany. He is currently an Assistant Professor with the Department of Computer Science, Government College University Faisalabad, Pakistan, where he has experience teaching for more than 12 years. His research interests include big data analytics, scientific computing, text classification, and bioinformatics.

MUHAMMAD UMER SARWAR received the Ph.D. degree from Government College University Faisalabad, Pakistan. He is currently an Assistant Professor with the Computer Science Department, Government College University Faisalabad. He has more than 18 year experience in teaching. He has several research publications in well-reputed journals. His research interests include natural language processing, data mining, database systems, and big data.