Tweet Credibility Detection for COVID-19 Tweets using Text and User Content Features

Vaishali Vaibhav Hirlekar[1]
Department of Computer Science and Engineering
Sir Padmapat Singhania University
Udaipur, Rajasthan, India

Arun Kumar[2]
Department of Computer Science and Engineering
Sir Padmapat Singhania University
Udaipur, Rajasthan, India

***Abstract*—The deadly COVID-19 pandemic is currently sweeping the globe, and millions of people have been exposed to false information about the disease, its remedies, prevention, and origins. During such perilous times, the propagation of fake news and misinformation can have serious implications, causing widespread panic and exacerbating the pandemic's threat. This increasing threat factor has given rise to considerable research challenges. This article is mainly concerned about fake news identification and experimentation is specifically performed considering COVID-19 fake news as a case study. Fake news is spread intentionally to mislead the people and therefore we need to identify user's involvement and it's correlation with additional features. The aim of this research is to develop a model that can predict the essence of a tweet given as an input with the help of multiple features. Our strategy is to make use of the tweet's text as well as the user's metadata and develops a model using natural processing technique and deep learning method. In this process, we have analyzed the behavior of the accounts, observed the impact of the various factors that can lead to fake news. The experimental analysis shows that hybrid model with text and content features have generated a benchmark result than the existing state of art techniques. We have obtained a best F1-score of 0.976 during the experimentation.**

***Keywords*—Fake news; machine learning; natural language processing; deep learning**

I. INTRODUCTION

Fake news and its consequences have the ability to impact a wide range of institutions, from a citizen's lifestyle to a country's international relations. Widespread use of social media sites been generating and exchanging more content than ever before, some of which is deceptive and has little bearing on fact. So far several related works has been carried out for collecting and identifying fake news, however no commercially viable system exists yet. Fake news isn't a new phenomenon but it has new repercussions and effects. Fake news can lead to the collapse and failure of the world's largest economies by mass exploitation, and it can be one of the most devastating "internet wildfires." Aside from political ramifications, fake news can and has resulted in personal defamation, distorted perspectives, and mass incitement on a variety of topics. It is much easier for the sources to produce this news than it is for people to embrace and share it. Fake news is the biggest danger to our ostensibly functioning democracy; in addition to distorting and corrupting ideologies, it has also resulted in real effects such as cyber defamation, cyber stalking and other cyber-attacks.

Detecting and identifying fake news on a social media is a difficult challenge. The rapid dissemination of false news has an impact on millions of people and their actual surroundings. The propagation of fake news is not a new issue on the social media sites [1]. Several firms and well-known individuals utilize various social media networks to promote their products and build their reputation. All of these operations persuade numerous people to share and enjoy the news. As a result of this process, fake news spreads over the web. In terms of a certain issue, the content, style, and media platform of fake news change with time and fake news tries to falsify linguistic information. Fake news may contain genuine evidence within a fabricated framework to promote a false assertion [2]. The term fake news has coined in the 2016 US Election primarily, which encouraged academicians and researchers to do the research in this direction [3]. The researchers tried to gather the data from various resources and then checked the actual authenticity of the news being spread. Since then people have been utilizing a variety of manual techniques to do the fact-checking, such as using fact-checking websites. These websites are crucial in spotting false news on the internet. There are a number of fact-checking and fake news identification research projects, methods, and applications available, most of which look at the issue from a veracity classification standpoint. Misinformation, disinformation, hoax, and rumor are all terms used in similar literature to describe fake news. The term "misinformation" refers to the spread of incorrect information without consideration of the real intention. The goal of misinformation is to fool the intended recipient of the information. Rumours or hoaxes are purposely crafted to appear accurate. The fake information is for the gullible. The person may not realize the actual authenticity and believes in on what is being spread through social media especially the social sites eager to increase their viewership.

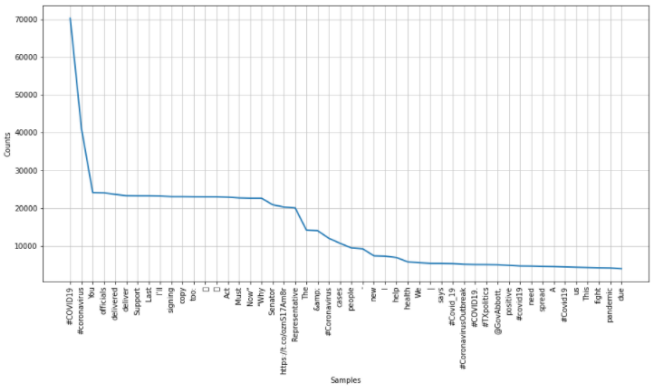
A. Covid 19 – An Infodemic

Covid 19 Infodemic has become more like a disease which is spreading rapid faster in the society through the dissemination of false information. Verifying the veracity, authenticity, and accuracy of given information is extremely difficult, especially when it concerns a horrifying disease that poses a threat to humankind. [4] COVID19, a virus that first appeared in Wuhan, China, in December 2019, has spread to 213 nations, regions, or territories throughout the world, resulting in roughly 3,478,418 fatalities as of May 24, 2021.

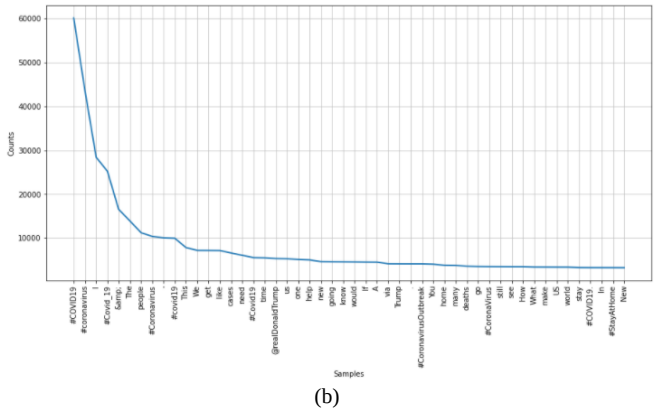
This infodemic posed a serious problem for public health alongwith social media channels including as Facebook, Instagram, WhatsApp and Twitter have become key sources of information on the crisis. In the COVID19 battle, fighting the infodemic is a new front. People trust the information that appeals to their emotions and personal opinions than information that is considered factual or objective in the 'posttruth' era. The following figure shows the news trend encountered during the Covid time. Fig. 1(a) shows plot of 50 most commonly used words in real tweet, whereas Fig. 1(b) shows plot of 50 most commonly used words in fake tweet.

The epidemic and infodemic elements of a pandemic are the two sides of the coin in today's highly digitalized society. This infodemics are usually fuelled by a combination of human and non-human system (bots), all of whom are pursuing essentially unknown aims. In this perspective, we present a methodology to detect fake tweets using the COVID-19 epidemic as a case study by using ensemble learning models with machine learning and Deep Learning Techniques.

In this research work, we have studied the differences between fake and real tweet based on behavioral, content based and comment based features of the tweet. The methodology implemented tackles the problems of fake tweet with the help of natural language processing toolkit and performs tweet text analysis as well. Different from the existing work, we take into consideration not only text characteristics, however also used user account characteristics for better results.



(a)



(b)

Fig. 1. (a) Shows Plot of 50 most commonly used Words in Real Tweet, (b) Shows Plot of 50 most commonly used Words in Fake Tweet.

The related work of fake news detection has been presented in Section II. Section III gives systematic overview of approaches used for Fake Tweet detection process. A Methodology along with model implementation has been discussed Section IV. The experimental setup and results are reviewed in Section V whereas Section VI makes the conclusion and recommendations for further research.

II. Related Work

Since all individuals communicate through a virtual world using social media, the methods of identifying fake information have been researched. Fake information spread by many social network users utilizing various platforms for financial or personal benefits. The evolution of social networking platforms encourages users to transmit and exchange information about their activities. However it has created a challenge for the researchers who want to secure user data from a variety of dangers. For researchers, detecting harmful information in the form of false news has become a significant challenge.

Looking back in time, misleading news isn't a new issue. Since a long time, there has been widespread anxiety about such news. The scientific community began to pay attention to the issue of fake information in the early 2000s, which expressed itself in the form of paid posters, review spam and rumor detections. The term "fake news," on the other hand, was popularized during the 2016 US Presidential Election [5] [6]. The impact of such news was unclear, although it spread some gossip, uncertainty, and deceit among users. Because of the growing media environment erroneous political information has also been circulating widely. The content of the news and the social circumstances are the most important factors in the fake news identification. News is classified into two categories: textual and visual, however emotions are also integral part of the news content. In addition, deep neural networks [7] [8] are used to frame latent textual representation.

Many of the researcher have also used different supervised and unsupervised, adversarial, user response methods. [9] Here author had created a false news detection algorithm utilizing machine learning approaches along with n-gram analysis. Here, author employed numerous characteristics collected using two distinct approaches and evaluated in six different machine learning settings. TF–IDF used as feature extraction method, and Support Vector Machine algorithm has been used in experimentation. In [10] researcher had developed a model to detect hoaxes or non-hoaxes distributed on social media platforms such as Facebook and created an automated fake news credibility inference algorithm to detect fake information. Here, [11] author had analyzed numerous variables such as user profile data and the relationship between users and the originator of the fake news. In [12] author, suggested a graph neural network-based technique and analyzed non-Euclidean data using a graph neural network. They often avoided certain written content by using unseen data for implementation. In addition, [13] author proposed model using twelve classifiers and evaluated on three datasets. The false prediction ratio of these ML classifiers is used to merge them. Based on their performance measures, Linear Support Vector Classifier, Logistic Regression and Passive Aggressive along with TF–IDF, CV, and HV feature extraction methods. In [14]

developed a volunteer-based crowd annotation tool by combining the perspectives of multiple stakeholders, the system were based on the Micro Mappers platform for English and Arabic tweets. To promote additional studies, [15] published consolidated content named CORD-19, which comprises 59,000 publications regarding COVID-19 and information associated with coronaviruses. The researchers have tried a variety of techniques to combat the COVID-19 Infodemic in recent months.

In [16] performed spatio-temporal analysis of the flow of information and the transmission of COVID-19. Author proposed a model [17] trained on several Indic Languages wherein fake news dataset tweets had performed better due to syntactic features. In Hindi, the model showed 79 percent and in Bengali 81 percent F-Score. [18] [19] conducted substantial research on the usage of machine learning strategies to resolve numerous COVID 19 difficulties. In [20] analyzed Facebook ads from 64 countries and discovered that about 5% of them included potential disinformation. However, none of these methods helped to address the disinformation issue by providing an explanation for the supplied fake assertion.

In [21] employed a natural language inference (NLI) model that was upgraded by adding internal semantic relatedness scores and ontological WordNet elements and performed claim verification on the FEVER Dataset. In [22] proposed a model developed using natural language processing methods and used different algorithms of machine learning, and deep learning that comprises of a categorization strategy that employs new twitter attributes. The approach is built in tandem with Apache Spark and achieved 79% accuracy using random forest algorithm. The author also noted that the emotion of tweets is essential in tweet categorization. In the process of detecting misleading information, [23] collected data from various fact-checking websites, performed filtering, preprocessing and feature selection operation on the text part of tweets and observed that Neural Network, Logistic Regression and Decision Tree classifiers has given the best performance from the different perspectives.

Here author used [24] two distinct methods to build a model that can implement constraint based task. To improve F1-score across several test sets by executing impact data purification with a high cleansing percentage (25%) and experimented a model with a 99 percent cleaning percentage and obtained the 54.33 percent F1 score and 61.10 percent accuracy. In this paper, author [25] had developed a BERT-based model with other important Twitter features. In addition, the method was extended to many Indian languages, and a mBERT-based model was used with Hindi and Bengali datasets and provided a methodology to solve the issue of data scarcity in low-resource languages. Using the annotated data, the model observed 81%, 79% percent F-Score in Bengali and Hindi Tweets, and 81 % F-score with zero shot model.

Although all of the aforementioned research indicated that studying and detecting fake news in the social media using various techniques is successful however several limitations were found. In some of the approaches various learning algorithm has been used to detect the fake news, which involves more processing time and brings limitation to the various accuracy parameters due to size of the dataset. Most of the technique concentrates only the text pattern of news, however there could be few factors that could differentiate between fake and real content.

In summary, fake content detection on any social media is a challenging task. A deep analysis is needed to identify user's involvement through various features. An exploratory study is performed in this article to explore the link between the content-based, comment-based, sentiment-based, and behavior-based characteristics, as well as the interrelationship between them. Furthermore, a hybrid model is proposed based on the integration of the text and metadata and tested on various standard deep learning and machine learning algorithm.

III. A Systematic Overview of Approach used for Fake Tweet Detection

The methodology implemented in this paper tackles the problems of fake tweet with the help of natural language processing toolkit and deep learning algorithms. Ultimate aim of the research is to classify each tweet into 2 distinct categories i.e. "real" or "fake". Experimentation has been carried out with the perspective wherein combination of user and content features will be used along with the tweet text. We have used multiple input models for the experimentation to handle continuous and numeric features efficiently.

A. Dataset used for the Experimentation

The dataset we selected contains the tweets of users from 29-03-2020 to 15-04-2020 using the following hashtags: #epitwitter, #coronavirusoutbreak, #covid19, #coronavirus, #ihavecorona, #corona, #coronavirusPandemic. From about 11 April 2020, the dataset also included the following additional hashtags: #StayHomeStaySafe, #TestTraceIsolate. The dataset contains variables as given: location, hashtag, title of the tweet, tweet text alongwith tweet account details. Dataset does not include retweets, although a count of retweets is provided as a variable. Alongwith the 'retweet_count' some of the other features are also included in the dataset i.e. favourites_count', 'followers_count', 'friends_count' which has been successfully used for improving the accuracy of the model. Around 303692 tweets have been used for the experimentation during the process, among which 156612 were fake tweets whereas 147080 were real tweets.

B. An Exploratory Analysis for Data Insight

An Exploratory Data Analysis has been performed for preliminary investigations on data in order to identify patterns, spot anomalies, testing hypotheses, and validating assumptions using summary statistics and graphical representations. The process of computationally identifying and classifying viewpoints in a text, with the goal of determining whether the writer has a +ve, -ve, or neutral attitude toward a certain topic, product, etc.

Fig. 2 shows Polarity and subjectivity score with sentiment and subjectivity flag wherein the subjectivity score range between [0.0, 1.0], polarity score shows in the range [-1.0, 1.0], wherein 1.0 being very subjective and 0.0 being highly objective.

	polarity	subjectivity	sentiment_flag	subjectivity_flag
0	0.000000	0.000000	neutral	objective
1	0.000000	0.000000	neutral	objective
2	0.541667	0.708333	positive	subjective
3	-0.100000	0.200000	neutral	objective
4	0.250000	0.333333	neutral	neutral

Fig. 2. Polarity and Subjectivity Score with Flag.

Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective sentence refers to factual information. Subjectivity is a float value which lies in the range of [0,1]. Here the value 0.7 represents that subjectivity is more, which ultimately refers that mostly it is a public opinion and not a factual information. Polarity, also known as orientation is the emotion expressed in the sentence. It can be positive, neagtive or neutral. Polarity is float value which lies in the range of [-1,1], here 0.5 refer to positive sentence.

Next distribution of user and content variables and relationships between other variables has been studied using pair plots in which the favorites, re-tweet, followers and friends count considered as a parameters.

Fig. 3 shows that favorites count and friends count are positively correlated. It also appears that re- tweet count and followers count has partial effect on favorites count. However during the analysis, all the features have been tested individually as well as in the combination to determine their impact on the tweet credibility detection. Furthermore, efficacy is also been tested with tweet content and user features in the deep learning and machine learning environment.

C. Approach used for Fake Tweet Detection

Extracting features from the content of the tweet is a reliable pattern recognition system and hence using natural language processing (NLP) is the most obvious solution to automatically identify fake news. To begin with this approach, data preprocessing has been carried out on tweet text part. Data preparation is a key stage in the modeling process, and the outcomes are dependent on how well the data has been preprocessed. In this approach, text normalization is performed which includes: converting numbers into words, converting letters, removing white spaces and punctuations, removing numbers and stop words. Pre-processed tweet text has been further given to the different vectorizer. For machine learning algorithms, TF-IDF is employed as it assigns a frequency score to words by emphasizing those that occur more often inside a document but not across documents. For deep learning algorithms, Glove has been used to obtain vector representations for words.

Tweet text feature is nothing but the pre-processed and normalized text through which the credibility of the content can be verified. Tweet text analysis done with NLP techniques has been able to evaluate the credibility of the tweet to some extent. However, text feature alone may not be enough to give better accuracy when it comes to detect the credibility. Features that come along with the tweet can be called as metadata of the tweet for exa. quote, favorites, retweet, followers, friends can play major role in the process of evaluation of the tweet because here tweets becomes more than strings with certain metadata added to them. This metadata becomes the features as an additional input dimension for an algorithm. With this perspective, we have tried to assess the tweet credibility by checking media content, account information and text characteristics. So here we have used 4 user features i.e. favourites_count', 'retweet_count', 'followers_count', 'friends_count' along with processing of tweet text that can make significance difference to the prediction of the news. Here we have tried to evolve a methodology to rate the credibility of the tweet based on the correlation with the additional features. In this phase, a tweet and user content feature has been used to perform the classification of the tweet.

- Behavioral Feature - considers user characteristics and linked user account attributes.
- Content based features - considers content of the tweet.
- Comment based features – considers the characteristics of the tweet itself.

The aim of our approach is how well multi-features-based method will be able to differentiate between the fake and real tweet. In this regards, we have tested different baseline algorithms like Random forest, logistic regression, Decision Tree, Naïve Bayes, XGBoost , Convolutional Neural Network , Bidirectional LSTM, and hybrid algorithm like CNN-BiLSTM for the tweet classification using the text and metadata features of the tweet. The methodology and algorithm have been discussed extensively in the next section that ultimately shows the flow of the process. Following tables are used to show features used for studies. Table I shows Tweet content features and Table II shows User content features used during the studies.

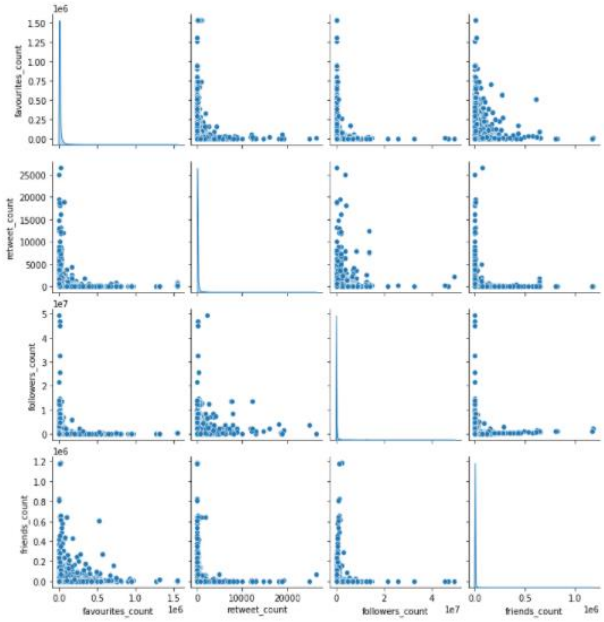


Fig. 3. Correlation between the Tweet Content Features.

TABLE I. TWEET CONTENT FEATURE USED FOR STUDY

Feature	Name	Type	Description
TC1	Title	Textual	Short text of the tweet, it is a summary of the topic's highlight
TC2	Text	Textual	Extended part of tweet that gives topic's details.
TC3	Source	Textual	Indicates source of the tweet.
TC4	is_quote	Boolean	A quote tweet is a re-tweet with some additional text attached. This parameter determines whether or not the selected tweet contains a quote.
TC5	is_retweet	Boolean	A retweet re-sends selected tweet. This field checks whether the selected tweet is a retweet.

TABLE II. USER CONTENT FEATURE USED FOR STUDY

Feature	Name	Type	Description
UC_1	favourites_count	Numerical	It's the number of tweets that given user has marked as favourite or in the account's lifetime, the number of Tweets this user has liked.
UC_2	Retweet_Count	Numerical	Retweet count always apply to the original tweet only, there is no counts for a "retweet" tweet, only the original, retweeted tweet. For example, if tweet B is a retweet of tweet A, and C is a retweet of B, in the end So in this example: B will have a count of 0, and so will C. A will have a retweet count of 2. As more individuals repost the tweet, this number may vary.
UC_3	*Followers_Count*	Numerical	No. of followers of the twitter account.
UC_4	*Friends_Count*	Numerical	It shows the number of friends of twitter account. However the ratio of followers to friends may well impart some useful information about the way in which the twitter account is being used.

IV. MODEL IMPLEMENTATION

A. Fake Tweet Detection with text and user Features using Statistical Approach

A statistical model is designed to derive the inference about the relationship between the variables and further used to predict the fake tweets. Based on the exploratory analysis, we have found the relationship between the user and content parameters, which are from different domain (i.e text and numeric). These features are combined further to build a model. In order to create a model that can handle continuous data and text data, following algorithm is used. Our dataset d has total of n data points: (d1, y1), (d2, y2), ……, (dn, yn) respectively , where di is the ith tweet and yi is its label. Each input sample, di comprises 2 input sub-sets here — tweet content feature (d^i_{TC}) and user content feature (d^i_{UC}).

Algorithm for Tweet and User features using Statistical Approach.

- Read tweet content feature (d^i_{TC}).
- nlp_input (V_{TC}) <- Process the d^i_{TC} data using.
- Preprocessing.
- Read all user content feature ($d^i_{UC1,UC2,UC3,UC4}$).
- tw=Convert to a matrix of TF-IDF features (V_{TC}).
- Concatenate the features (tw, UC1, UC2, UC3, UC4).
- Spliting into Train and Test set.
- Model <- classifier ( ).
- Model.fit (features_train data).
- Prediction=Model (features_test data).

Experimental model 1 used Machine learning approach with combination of text and user feature which consists of Text Pre-processing, Tokenization, TF-IDF vectorizer for text part and then converted all the features generated from metadata and n-gram frequencies of the text into a matrix. Each row represents a tweet and each column the value of one of user features. Further we have used different classifier algorithms like Random Forest, Naïve Bayes, Logistic Regression, Naive Bayes, Decision Tree and XGBoost for the experimentation.

B. Fake Tweet Detection with text and user Features using CRED_Tweet Model

In this approach, algorithm operates in two phases. In the first phase extraction, d^i_{TC} data i.e. text input is processed using regular expression with the help of nltk library to perform basic data cleaning operation and then pre-trained glove model is used to convert text into embedding. In the next phase, User features d^i_{UC} are then interpolated to higher dimensional dense feature vectors termed meta_input d^i_{UC} through separate convolution kernel. In the proposed (CRED_Tweet) model shown in Fig. 4, a combination of multiple inputs CNN-BiLSTM architecture is successfully used providing text and metadata information through different convolution layers.

Here, vocabulary size V is used to represent the tweet text of length N. The embedded input is passed through 3 different convolution layer Conv1D (128, 5, activation='relu') followed by max pooling layer. Here, in the convolution layer used 128 filters. The convoluted vectors is been given to Bidirectional LSTM layer with 128 internal unit with dropout=0.3 and kernel_regularizer attributes and further to flatten. In the next phase, User features d^i_{UC} are then interpolated to higher dimensional dense feature vectors termed meta_input d^i_{UC} through separate convolution, pooling and flatten layers and further concatenated both the dense tensors, $Z_{i\,(ht)} \oplus X_i$'. The vector Z_i is then passed across a fully connected network, with the probability distribution across the two classes being regularized using a dropout layer.

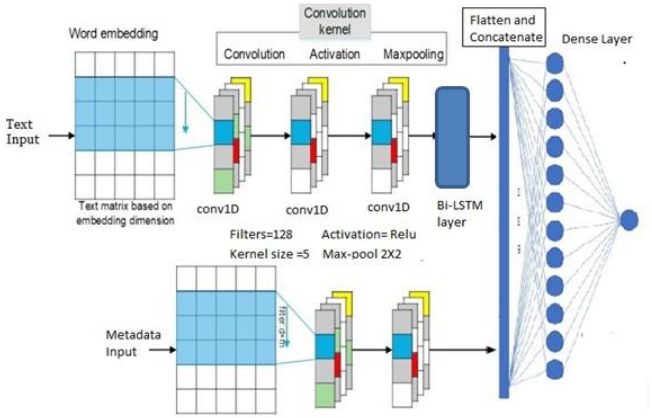


Fig. 4. Block Diagram of CRED_Tweet Model using user and Content Features.

Phase I: Extracting text features

In this model, tweet text features (TC) and User content feature (UC) are based on a variant of CNN and processed individually through two different CNN model as the nature of both the feature are different i.e. text and numerical respectively. Although CNNs are mostly employed in image classification [26] or object identification, they have also demonstrated noteworthy performance in several Natural Language Processing (NLP) applications including text classification [27] [28]. Using the convolutional approach, the neural network produce local features around each word of the neighbouring word in the first phase and later they are combined using a max operation. As a result, we use CNN to model textual characteristics for the identification of fake news. Fig. 4 shows the layered processing of the CRED_Tweet model using user and content features.

Let Z be the text input, here maximum length of the tweet considered as n and padded with sequence length m. $\in d^{i}_{TC.}$

$$Z_i^{TC} = Z_{i,1} \oplus Z_{i,2} \oplus Z_{i,3} \oplus .Z_{i,n} \quad (1)$$

Here, each tweet is been re-presented as a matrix and then convolution filter are applied to derive new features. Here, we have used convolutional filters $w \in R^{Hl \times Wl \times Dl \times N}$ to construct the new features. Equation (2) shows default convolution function. When the word matrix is been processed through a convolution layer, it can produce the feature Z i.

$$Z = W^{T}.X + b \quad (2)$$

$$Z_i = f\left(W^{T}. Z_i^{TC} + b \right) \quad (3)$$

Here the b represents the bias, and · is the convolutional operation. A layer or convolution kernel includes convolution layer, activation layer and pooling layer as shown in Fig. 4. The function f is the non-linear transformation and also includes ReLU activation layer here.

$$Z_i = \max\{0, Z_i\} \quad (4)$$

During the experimentation, Zi has been passed through 3 different convolution layers of 128 neurons and having filter size = 5 and activation='relu. The filter generates a feature map by running over every potential window of words in the tweets. A max-pooling layer [24] is utilized to get the maximum feature map. The maximum value is denoted as Z_i'.

$$Z_i' = \max\{ Z_i \} \quad (5)$$

By saving the most significant convolutional findings for false news detection, the max-pooling layer can considerably increase the model's robustness. The CNN has the advantage over the LSTM as it decreases the number of dimensions in the input features that must be provided to a sentiment classifier or a natural inference prediction model after the feature extraction stage.

These token vectors (Z_i') are further encoded using a Bi-LSTM, using the forward and backward layers which processes the N vectors in opposite directions. a hidden state hft is emitted by the forward LSTM at each time-step, which is concatenated with the corresponding hidden state hbt of the backward LSTM to produce a vector $ht \in R^{Hl \times Wl \times Dl \times N}$.

$$Z_{i\,(ht)} = Z_{i\,(hft \oplus hbt)} \quad (6)$$

Phase II: Extracting metadata features

Let X be the metadata numeric input $\in d^{i}_{UC}$. X_i^{UC} has been passed through 2 different convolution layers of 128 neurons and having filter size = 5 and activation='relu.

$$X_i^{UC} = X_{i,1} \oplus X_{i,2} \oplus X_{i,3} \oplus \dots X_{i,n} \quad (7)$$

$$X_i = f\left(W^{T}. X_i^{UC} + b \right) \quad (8)$$

$$X_i = \max\{0, X_i\} \quad (9)$$

$$X_i' = \max\{ X_i \} \quad (10)$$

At the end, concatenate both the features (text and metadata) feature.

$$Z_i^{'} = Z_{i\,(ht)} \oplus X_i' \quad (11)$$

A convolutional network's fully connected layers are essentially a multilayer perceptron that used to map the $m^{(l-1)}{}_1 \times m^{(l-1)}{}_2 \times m^{(l-1)}{}_3$ activation volume from the preceding various layers into a class probability distribution. As a result, the multilayer perceptron's output layer will contain $m^{(l-1)}{}_1$ output neurons, where i specifies the number of layers in the multilayer perceptron.

$$y_i = f\left(Z_i^{'}\right) \quad (12)$$

Pred (y for given di TC, di UC ; θ) = activation function(f) on (Zi').

Here, Z_i' signifies the changed vector after passing through the relevant feed forward sub-network and sigmoid activation function, while θ denotes the model parameters employed throughout the experiment.

Loss function and optimizer:

The aim of any optimization problem is to minimize the cost function, which means of measuring how accurate the data is. We utilized Binary cross entropy in this case, which compares each of the predicted probabilities to the actual class output, which can be 0 or 1. The score is then calculated, penalizing the probabilities depending on their deviation from the predicted value. This refers to how close or far the value is

to the real value. The negative average of the log of corrected projected probability is shown by Binary Cross.

Entropy.Loss= abs(y_pred – y_actual) (13)

Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	[(None, 1000)]	0	
embedding_1 (Embedding)	(None, 1000, 100)	72004300	input_2[0][0]
conv1d_5 (Conv1D)	(None, 996, 128)	64128	embedding_1[0][0]
max_pooling1d_5 (MaxPooling1D)	(None, 199, 128)	0	conv1d_5[0][0]
conv1d_6 (Conv1D)	(None, 195, 128)	82048	max_pooling1d_5[0][0]
metadata_input (InputLayer)	[(None, 4, 1)]	0	
max_pooling1d_6 (MaxPooling1D)	(None, 39, 128)	0	conv1d_6[0][0]
conv1d_8 (Conv1D)	(None, 4, 128)	384	metadata_input[0][0]
conv1d_7 (Conv1D)	(None, 35, 128)	82048	max_pooling1d_6[0][0]
max_pooling1d_8 (MaxPooling1D)	(None, 2, 128)	0	conv1d_8[0][0]
max_pooling1d_7 (MaxPooling1D)	(None, 1, 128)	0	conv1d_7[0][0]
conv1d_9 (Conv1D)	(None, 2, 128)	32896	max_pooling1d_8[0][0]
bidirectional_1 (Bidirectional)	(None, 256)	263168	max_pooling1d_7[0][0]
max_pooling1d_9 (MaxPooling1D)	(None, 1, 128)	0	conv1d_9[0][0]
flatten_2 (Flatten)	(None, 256)	0	bidirectional_1[0][0]
flatten_3 (Flatten)	(None, 128)	0	max_pooling1d_9[0][0]
concatenate_1 (Concatenate)	(None, 384)	0	flatten_2[0][0] flatten_3[0][0]
dense_2 (Dense)	(None, 128)	49280	concatenate_1[0][0]
dense_3 (Dense)	(None, 2)	258	dense_2[0][0]

Total params: 72,578,510
Trainable params: 72,578,510
Non-trainable params: 0

Fig. 5. Fitting the Convolutional Neural Network with Bilstm Model.

Optimizers are techniques or strategies for changing the characteristics of a neural network, such as weights and learning rate, to minimize losses. To minimize losses, the optimizer determines how to alter the weights or learning rates of the neural network. Adam optimizer has proven benchmark outcomes above existing state-of-the-art algorithms by training the neural network in less time and more effectively. Fig. 5 shows fitting of the convolutional neural network with Bi-LSTM model. From the given figure, we can analyse how multiple inputs of different types are processed here through different layers. Continuous input i.e. text or content features of the tweets has been processed through convolutional layers and then passed through Bi-LSTM layer whereas metadata features i.e. user features has been processed through separate convolution layer, then concatenated output of both layer and provided further to fully connected layer for final predictions. In this process, total trainable parameters encountered are 72,578,510.

Precision: Conversely, precision score represents the ratio of true positives to all events predicted as true. In our case, precision shows the number of articles that are marked as true out of all the positively predicted (true) articles:

Precision = TP/ (TP+FP) (14)

Recall: Recall represents the total number of positive classifications out of true class. In our case, it represents the number of articles predicted as true out of the total number of true articles.

Recall = TP/(TP+FN) (15)

F1-Score: F1-score represents the trade-off between precision and recall. It calculates the harmonic mean between each of the two. Thus, it takes both the false positive and the false negative observations into account. F1-score can be calculated using the following formula:

F1= 2. (Precision.Recall)/(Precision+Recall) (16)

Accuracy: Accuracy is often the most used metric representing the percentage of correctly predicted observations, either true or false. To calculate the accuracy of a model performance, the following equation can be used:

Accuracy = TP+TN / (TP+TN+FP+FN) (17)

The predicted results are evaluated with confusion matrix and other measures like True negative rate (Specificity), True positive rate (TPR), Precision, Recall (Sensitivity), F1-score, accuracy, PRC (Precision-Recall curve) and ROC (Receiver operating curve) etc. Tables III and IV shows the performance results of various parameters. Performance of several user profile categories, tweet content elements, and a combination of both has been examined. The best accuracy found is with decision tree 92.56 among all the ML algorithms along with 92.13 % precision and 92.54 % recall. However the further experimentation with deep learning approaches had shown a benchmark result over the existing state of art techniques.

Deep learning-based analysis has a higher accuracy and detection rate than machine learning. In the experiment 2, CNN with Bi-LSTM is used with the embedding layer and convolution layers with 128 tensors, 5 filters and with relu activation. Further Bi-LSTM is used with 128 neurons and 0.3 dropout and recurrent dropout, with regularizers.l2 (0.01). User features are also separately processed through different CNN layer with with 128 tensors, 5 filters and with relu activation. ‘Relu’ and ‘softmax’ activation function used in dense layer. Model is further compiled with ‘binary_ cross entropy’ loss and ‘adam’ optimizer and found best accuracy with 97.60 with text and metadata feature as shown in Table IV. Size of train set, test and validation set is as given here: 242953, 151846, 151846. The result shown in Table IV, deep learning scenarios, shows that combining certain features is based on user content and tweet content improves accuracy. Among all the algorithms, CRED_Tweet Approach had achieved 98.44 % precision, 96.56 % recall and 97.60 % accuracy.

TABLE III. SHOWS ACCURACY PARAMETERS FOR DIFFERENT MACHINE LEARNING CLASSIFIERS WITH TEXT AND METADATA FEATURES

Measures in %	**LR**	**NB**	**RF**	**DT**	**XG Boost**
Precision	85.17	91.80	91.72	92.13	90.91
Recall	91.02	74.40	90.97	92.54	91.41
F1 score	87.98	84.15	91.64	92.55	91.41
Accuracy	87.98	84.39	91.65	92.56	91.42

TABLE IV. SHOWS ACCURACY PARAMETERS FOR DIFFERENT DEEP LEARNING CLASSIFIERS WITH TEXT AND METADATA FEATURES

Measures in %	**Bi-LSTM**	**CNN**	**CRED_Tweet Approach**
Precision	87.22	98.41	98.44
Recall	93.02	95.44	96.56
F1 score	90.01	97.04	97.60
Accuracy	90.01	97.05	97.60

C. Analysis of the Result

The performance of the statistical ML models has been observed using ROC and precision-recall curve. ROC curves summaries the trade-off between the true positive rate and the false positive rate for a predictive model with varying probability thresholds whereas Precision-Recall curves illustrate the trade-off between a predictive model's actual positive rate and positive predictive value when different probability thresholds are used. The integral or an estimate of the area under the precision-recall curve is summarized as AUC (Area under curve). Fig. 6 illustrates ROC-AUCs and PRC-AUCs for the deployed models. Here, Random Forest and XG Boost Algorithm gives 0.97 %, Naïve Bayes gives 0.83 %, Logistic Regression and Decision Tree gives 0.93 % ROC-area under curve. The precision-recall curve is constructed by calculating and plotting the precision against the recall for different classifiers at a variety of thresholds. PRC identifies the Positive Predictive Value (precision) for each corresponding value on the sensitivity (recall) scale. Here, Random Forest and XG Boost Algorithm gives 0.97 %, Naïve Bayes gives 0.81 %, Logistic Regression and Decision Tree gives 0.94 % and 0.89 % respectively area under curve for PRC. In our experimentation, Decision Tree had the best precision and recall i.e. 92.13% and 92.54 % with the 92.56 % accuracy which is best among all the algorithms.

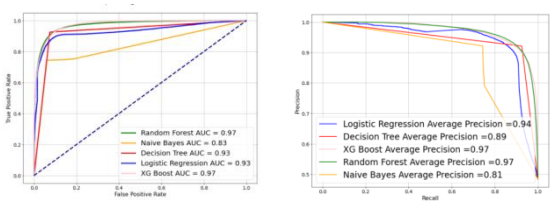


Fig. 6. Shows ROC and Precision-Recall Curve for LR,DT,RF,NB and XG Boost Classifier with Text and Metadata.

During the analysis, all the metadata features have been tested individually as well as in the combination to determine their impact on the tweet credibility detection. However, it has been observed that combination of both news content and user content features improves detection rate. The proposed hybrid model CRED_Tweet trained in order to improve work in this domain. The model outperforms LSTM with similar weights and shorter training time in terms of test accuracy. As a result, quicker training with CNN is feasible, decreasing the training time required for big datasets. Fig. 7 shows the plots of the training and test accuracy and loss values of the model over the 05 epochs. Model loss figure shows good fit learning curve which shows that training and testing loss that decreases to a point of stability remains almost the same in all epochs.

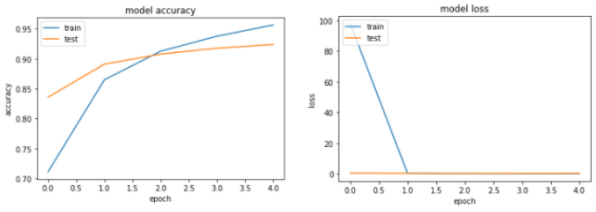


Fig. 7. Shows Accuracy and Loss of CRED_Tweet Model.

D. Comparative Analysis

We compare the CRED Tweet model to a few state-of-the-art approaches, which are listed below. TI-CNN technique used by author [29] with text and picture for fake news detection is by combining implicit and explicit characteristics. This technique has used 8,074 real news and 11,941 fake news which gave 92.2 % precision, 92.7 % Recall, 92.10 % F1-score and accuracy. In [30], author have used machine learning approaches with multiple features extracted from different sources, which used 2282 Buzz Feed news related to US election and found 85% AUC with Random Forest, 80 % with KNN and 86 % with XGB. [13] Here author have used Multilingual Approach for Fake Tweet Detection. For Indic i.e. Bengali and Hindi Languages & English give 92.75% Precision, 62.95% Recall, 75% F1-score and 81% accuracy. In [22] various machine learning algorithm has been experimented on Covid-19 epidemic fake news dataset which contains 5000 real tweets and 5000 fake tweets. The method gives 85 % precision, 82% F1-score and 79 % accuracy. Table V and Fig. 8 shows comparative evaluation with other approaches w.r.t precision, recall, F1-score and accuracy parameter.

TABLE V. SHOWS COMPARATIVE EVALUATION OF CRED TWEET WITH OTHER APPROACHES

Author	Methodology	Dataset	Results	Limitation
Yang et al. 2018 [29]	TI-CNN technique to evaluate picture and text for fake news analysis by combining explicit and latent characteristics..	20,015 news, i.e., 11,941 fake news and 8,074 real news	Precision-92.2 Recall-92.7 F1-score-92.10	The model trained only on CNN which may work better with picture but for text RNN model is needed.
Reis et al. 2019 [30]	Machine learning approaches are used with multiple features extracted from different sources	2282 Buzz Feed news related to US election	RF-85% KNN - 80% SVM - 79%	Accuracy for detecting fake account is very low due to small dataset.
D. Car et al. 2020 [25]	Multi-Indic-Lingual Approach used for COVID Fake-Tweet Detection	COVID-19 multilingual tweet dataset for Indic Languages (Hindi and Bengali) & English	Precision-92.75 Recall-62.95 F1-score-75.00	Accuracy for detecting fake account is very low due to small dataset.
Y. Madani et al. 2021 [22]	The various machine learning algorithm has been experimented on Covid-19 epidemic fake news dataset	FakeNewsNet dataset contains 10,000 fake and real tweets (5000 fake tweets and 5000 Real tweets).	RF-79%, DT-62% LR-60%, SVM-72% NB -53%, MLP-48 %	Accuracy for detecting fake account is very low due to small dataset.

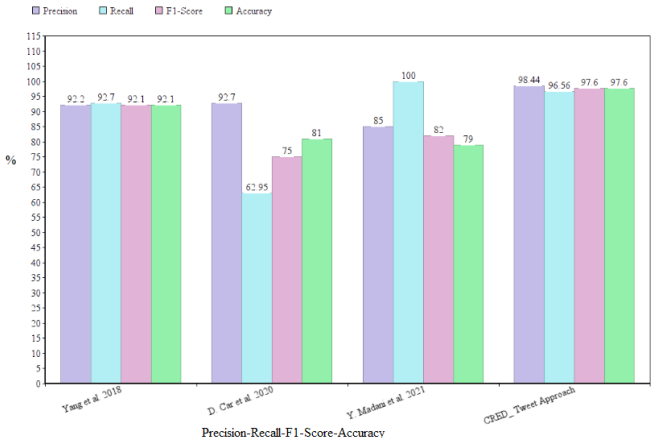


Fig. 8. Comparative Evaluation with other Approaches.

With respect to all the approaches specified above, CRED Tweet Approach uses CNN - Bi-LSTM for Tweet content feature and CNN model for user content features. In this approach 303692 tweets has been used for the experimentation during the process, among which 156612 were fake tweets whereas 147080 were real tweets. CRED_Tweet Approach had achieved 98.44 % precision, 96.56 % recall and 97.60 % accuracy which is considerably better, also the size of the dataset used during the experimentation is quite large as compared to other state-of-the-art approaches.

The methodology implemented in this paper tackles the problems of fake tweet with the help of natural language processing toolkit and deep learning algorithms. Ultimate aim of the research is to classify each tweet into two distinct categories i.e. “real” or “fake”. Experimentation has been carried out with the perspective wherein combination of user and content features will be used along with the tweet text. We have used multiple input models for the experimentation to handle continuous and numeric features efficiently.

V. Conclusion

In this research work, we have studied the differences between fake and real tweet based on behavioral, content based and comment based features of the tweet and further help to classify fake tweets in an extremely dedicated domain of COVID-19.

The methodology implemented tackles the problems of fake tweet with the help of natural language processing toolkit and performs tweet text analysis as well. Different from the existing work, we take into consideration not only text characteristics, however also used user account characteristics for better results and we have found that the efficacy of fake tweet detection is improved using tweet content features and user content features.

Our proposed model outperforms better than other baseline deep learning and machine learning approaches. Overall, the use of ANN in the identification of fake news appears to be promising. Aside from CNN and Bi-LSTM, we'll look at more complex neural network architectures in the future. When traditional models are combined with task-specific function engineering techniques, they can be extremely useful. In future, we aimed at doing in-depth exploratory analysis on the tweet in order to find out the indirect features that can affect the credibility of the news. Despite the enormous amount of existing works on fake news identification and detection, there is still room for improvements, and new profound developments into the nature of fake news can lead to more effective and accurate models.

References

[1] S. Sahoo and B.B. Gupta. Multiple features based approach for automatic fake news detection on social networks using deep learning. Applied Soft Computing, Volume 100, 106983, 2021. doi:10.1016/j.asoc.2020.106983.

[2] S. Feng, R. Banerjee and Y. Choi. Syntactic stylometry for deception detection. In Proc. of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers – Volume 2(ACL '12), Pages 171–175, 2012. doi/10.5555/2390665.2390708.

[3] B. Horne, S. Adali. This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News, 2017. ArXiv, abs/1703.09398.

[4] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C.M. Valensise, E. Brugnoli, A. Schmidt, P. Zola, F. Zollo and A. Scala. The COVID-19 social media infodemic. Sci Rep 10, 16598 (2020). https://doi.org/10.1038/s41598-020-73510-5.

[5] E. Bakshy, S. Messing, L. A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. SCIENCE05 JUN 2015 : 1130-1132. DOI:10.1126/science.aaa1160.

[6] M. Barthel, A. Mitchell and J. Holcomb. Many Americans believe fake news is sowing confusion, Pew Res. Center 15 (12) (2016).

[7] H. Karimi, P. Roy, S. Sadiya and J. Tang. In Proceedings of the 27th International Conference on Computational Linguistics, Pages 1546—1557, 2018. https://aclanthology.org/C18-1131.

[8] S. Hosseinimotlagh and E. E. Papalexakis, Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. In the Proc. of Misinformation and Misbehavior Mining on the Web Workshop held in conjunction with WSDM , 2018.

[9] H. Ahmed, I. Traoré and S. Saad. Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. ISDDC 2017. DOI:10.1007/978-3-319-69155-8_9.

[10] E. Tacchini, G. Ballarin, M.L. Vedova, S. Moret and L.D. Alfaro (2017). Some Like it Hoax: Automated Fake News Detection in Social Networks. ArXiv, abs/1704.07506.

[11] J. Zhang, B. Dong and P. S. Yu. Fake news detection with deep diffusive network model, 2018. arXiv:1805.08751.

[12] Y. Han, S. Karunasekera and C. Leckie, Graph neural networks with continual learning for fake news detection from social media, 2020. arXiv:2007.03316.

[13] D. Kar, M. Bhardwaj, S. Samanta, A. Azad. No Rumours Please! A Multi-Indic-Lingual Approach for COVID Fake-Tweet Detection, 2020. arXiv:2010.06906.

[14] F. Alam, F. Dalvi, S. Shaar, N. Durrani, H. Mubarak, A. Nikolov, G. Martino, A. Abdelali, H. Sajjad, K. Darwish and P. Nakov. Fighting the covid-19 infodemic in social media: A holistic perspective and a call to arms,2020. arXiv:2007.07996 [cs.IR].

[15] L. Wang, K. Lo, Y. Chandrasekhar, R. Reas, J. Yang, D. Eide, K. Funk, R. Kinney, Z. Liu, W. Merrill, P. Mooney, D. Murdick, D. Rishi, J. Sheehan, Z. Shen, B. Stilson, A.D Wade, K. Wang, C. Wilhelm, B. Xie, D. Raymond, D.S. Weld, O. Etzioni and S. Kohlmeier. Cord-19: The covid-19 open research dataset. ArXiv (2020).

[16] L. Singh, S. Bansal, L. Bode, C. Budak, G. Chi, K. Kawintiranon, C. Padden, R. Vanarsdall, E. Vraga and Y. Wang. A first look at covid-19 information and misinformation sharing on twitter,2020. arXiv preprint arXiv:2003.13907.

[17] F. Alam, F. Dalvi, S. Shaar, N. Durrani, H. Mubarak, A. Nikolov, G. Martino, A. Abdelali, H. Sajjad, K. Darwish and P. Nakov. Fighting the covid-19 infodemic in social media: A holistic perspective and a call to arms,2020. arXiv:2007.07996 [cs.IR].

[18] W. Naude 2020. Artificial intelligence against covid-19: An early review. https://towardsdatascience.com/artificial-intelligence-against-covid-19-an-early-review-92a8360edaba.

[19] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam and M. Luengo-Oroz. Mapping the landscape of artificial intelligence applications against covid-19, 2020.arXiv preprint arXiv:2003.11336.

[20] Y. Mejova , I. Weber and L.Fernandez-Luque . Online health monitoring using facebook advertisement audience estimates in the united states: evaluation study. JMIR public health and surveillance, 2018.doi:10.2196/publichealth.7217.

[21] Y. Nie, H. Chen and M. Bansal. Combining fact extraction and verification with neural semantic matching networks. In Proc. of the AAAI Conference on Artificial Intelligence, volume 33, pages 6859–6866, 2019. doi: 10.1609/aaai.v33i01.33016859.

[22] Y. Madani, M. Erritali, B. Bouikhalene. Using artificial intelligence techniques for detecting Covid-19 epidemic fake news in Moroccan tweets. Results in Physics, Volume 25, 2021. doi:10.1016/j.rinp.2021.104266.

[23] M. K. Elhadad , K. F. Li And F. Gebali. Detecting Misleading Information On Covid-19. In proc. of IEEE Access, vol. 8, pp. 165201-165215, 2020, doi: 10.1109/ACCESS.2020.3022867.

[24] Y. Bang, E. Ishii, S. Cahyawijaya, Z. Ji and P. Fung. Model Generalization on COVID-19 Fake News Detection, 2021.doi:arXiv:2101.03841.

[25] D. Kar, M. Bhardwaj, S. Samanta, A. Azad. No Rumours Please! A Multi-Indic-Lingual Approach for COVID Fake-Tweet Detection, 2020. arXiv:2010.06906.

[26] A. Krizhevsky, I. Sutskever and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Proc. of the 25th International Conference on Neural Information Processing Systems - Volume 1, Pages 1097–1105, 2012. doi: 10.5555/2999134.2999257.

[27] D. Zeng, K. Liu, S. Lai, G. Zhou, J. Zhao. Relation Classification via Convolutional Deep Neural Network. In Proc. of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages- 2335–2344, 2014. https://aclanthology.org/C14-1220.

[28] Y. Kim. Convolutional neural networks for sentence classification. In Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages- 1408.5882, 2014. doi:10.3115/v1/D14-1181.

[29] Y. Yang, L. Zheng, J. Zhang, Q. Cui, Z. Li and P. S. Yu. Convolutional Neural Networks for Fake News Detection. 2018. DOI: arXiv:1806.00749.

[30] J. C. S. Reis, A.Correia, F. Murai, A. Veloso and F. Benevenuto. "Supervised Learning for Fake News Detection," in IEEE Intelligent Systems, vol. 34, no. 2, pp. 76-81, March-April 2019, doi: 10.1109/MIS.2019.2899143.