IJICTR
International Journal of Information & Communication Technology Research

Volume 12- Number 4 – 2020 (46 -59)

Fake News Detection Based on Social Features by Ordered Weighted Averaging Fusion

Mehdi Salkhordeh Haghighi*
Faculty of Computer Engineering and IT
Sadjad University
Mashhad, Iran
haghighi@sadjad.ac.ir

Nasim Eshaghian
Faculty of Computer Engineering and IT
Sadjad University
Mashhad, Iran
Na.eshaghian454@sadjad.ac.ir

Received: 3 October 2020 - Accepted: 5 December 2020

Abstract—**Today, different groups of people use social media in their businesses and normal daily activities specially for accessing news and their favorite information in various fields. Facing with huge amounts of information and news in social media makes different challenges for the users. One of the main challenges of the users is distinguishing valid news and information from invalid and fake ones. Fake news means low quality news containing inaccurate or invalid information. Because of the fast and widely spread of the news in social media, they may have very destructive effects on the user's social behavior. Therefore, the fake news should be identified and banned as soon as possible. To overcome the challenge of identifying fake news, in this manuscript a method is introduced to use profile features of the users and some features of the tweets in twitter to determine the possibility of a tweet being fake. This method also uses ordered weighted averaging as a data fusion method to increase the accuracy of the detection. To determine the effectiveness of the presented method, some experiments are designed based on the known datasets from twitter. The evaluations of the results of these experiments indicate effectiveness of the proposed method.**

Keywords— fake news detection; Data fusion; social features; tweet features; user profile features; OWA

I. INTRODUCTION

Nowadays, people spend most of their time in social media. Increasing growth of social media usage has huge impact on different aspects of people's lives. many researches have been done in social media on people's behavior in communities [1] [2]. Because of high speed, low cost, and ease of access to information, people are increasingly using social media. For example, in December 2016, the PEW research center announced that approximately 62% of adults follow news from social media. However, in 2012, only 46% of people received this information [3]. On the other hand, poor quality of news in social media is their major drawback. In fact, fake news is the news that contain a lot of inaccurate

* Corresponding Author

or invalid information [4]. The rise of social media has been accompanied by a sharp increase in the spread of fake news issues among people. In other words, fake news is the misleading news that include fake contents in the form of serious expressions, lies and humor [5].

The importance of recognizing fake news in social media is because of the negative effects of wide spreading fake news or rumor [6]. These effects overshadow both the individuals and the communities [7]. Some of the drawbacks of fake news are destroying the balance of news ecosystem, encouraging the reader to believe invalid news and misleading the reader about the truth of the news [3]. Therefore, it is necessary to provide a useful method to detect fake news in social media to reduce damaging effects of them. In order to identify fake news, it is necessary to classify all the news based on their accuracy and reliability [8].

Detecting fake news in social media has special challenging features. These features are used by the writer to write the message precisely to mislead the reader so that making it difficult to detect the fake news from the content of the message. Various empirical studies have been conducted in Facebook usage among adolescent girls. These studies have consistently found that the visual and interactive aspects of the platform have the greatest influence on body image issues [9]. Despite this, highly visual social media (HVSM) such as Instagram have yet to be robustly researched. That is why we need a deeper study about all the news produced in social media. One of the key points that should be considered in fake news detection is the relationship between user profiles and fake news [10]. For example, if a verified user with a large number of followers talks about a news that may be fake, it is most likely fake. Most of the presented methods for detecting fake news use the features extracted from the contents of the message instead of other features of news such as social context [10].

Therefore, in the presented method in this manuscript, social features are also considered for two main reasons. The first is that the writer of the fake news exactly intends to mislead the reader, hence he/she tries to change the linguistic and content features so that the truthfulness of the news is not recognizable. Therefore, relying just on the content features is not a rational task for identifying fake news. The second is that social features of news reflect the characteristics of the activities in the social environment in which the news is published. [11] These features are user social features and indicate how the user interacts online (by following others and responding to each other's messages) [12].

Based on these reasons, we assume that social features have positive impact on fake news detection in social media. Social features are divided into two main categories: user and message features. User features are extracted from the user profile, for example, the number of followers, the number of friends and etc. Message features are extracted from the message contents and indicate how the users react to the message.

Recently, machine learning based methods [6] are increasingly being applied to the fake news detection. These methods focus on selecting some features and incorporating these features on classifiers such as support vector machines (SVM), k nearest neighbors (KNN), etc. Building a complex model based on some simple components is effective to improve performance of fake news detection [13]. Such a model uses different simple base classifiers to decide about the strength of the news being fake. Then, to get a better result, a data fusion method combines the output results of the base classifers. The fusion method highly affects the accuracy of the fake news detection. However, data fusion approaches have not yet been explored for fake news detection explicitly. On the other hand, some very complex methods based on convolutional neural networks are also introduced for fake news detection. But, these methods have very complex structures and need large amounts of traininh data [14].

In this manuscript, we use social features to train the base classifiers and then by using a data fusion approach, final decision is made to increase the accuracy of detecting fake news. Therefore, the main contribution in this manuscript is introducing a data fusion approach and using social features for training some classifiers used by the fusion method to increase the accuracy of detecting fake news. Moreover, by defining a fuzzy metric for fake news detection and using a threshold to distinguish fake news from the others, a more reliable system is made that is adaptable with different environments. On the other hand, one of the strength of the proposed method is that it uses simple base classifiers that may be trained more easily with lower amounts of training data than the complex CNN based methods.

Therefore, the main reasons that motivated the authors to focus on detecting fake news in social media are : (1) widely spread of using social media by different types of users, (2) increasing the number of users that produce fake news to mislead the readers of the news, and (3) existing complex systems need large amounts of time and resources to detect fake news, but, using simple elements (base classifiers) and a fusion method reduces the complexity of the detection method while needs lower amounts of training data than the others. As a result, themain contributions are summarized as follow:

- Combining message features with the features in the user profile for fake news detection in twitter.
- Using ordered weighted averaging (OWA) as a fusion method and using a method to compute the weights.

The structure of the remaining parts of this manuscript are as follow. In section2, some related works are summarized and briefly described to indica te the ways that the fake news detection methods have been used. In section 3, the proposed method for fake news detection is presented. And described. In section 4, some experiments are designed to compare the proposed method with some other known methods. Finally, in section 5, the concusion is made and some points are presented for future works.

II. Related Works

In this section, some of the most known previous studies on detecting fake news has been described briefly. Different approaches have been used recently to detect fake news in social media. One of the categorizations of the methods are based on the features they consider. In this categorization, two types of features are used: content features and context features. Content features are extracted from the body of the message. On the other hand, context features are extracted from the user profile and some related features [15].

Another categorization is based on the method used to detect fake news. In this category, the two general methods are machine learning and deep learning. The machine learning methods basically are the classifiers that classify the news as fake and non-fake. In machine learning methods, some data fusion techniques are also used to reduce complexity of the classifiers and increase the accuracy. Some research in each of these categories are briefly described in this section.

A. Fake news detection based on content features

In this part, content-based approaches have been discussed. These methods utilize textual features such as writing style features, word vectors, part-of-speech tags, question marks, exclamation marks, capital letters, sentiments, emotion features, manipulation features, grammatical features, and readability features [16].

Potthast et al. [17] discussed about linguistic feature such as quotes count, external links count, paragraph count, and average paragraph length. They also have proposed an unmasking method which recognizes the depth of difference between two messages in terms of writing style by using a random forest method. They indicated that how a style analysis can detect fake news on BuzzFeed dataset.

Singh et al. [18] have explored several conventional classifiers including Logistic Regression, Linear discriminant analysis, Quadratic discriminant analysis, K Nearest Neighbors, Naïve Bayes, SVM, CART (Classification and Regression Tree), and Random Forest. Feature extracted in this article was a combination of text and visual features, which include organization features (such as word count, words per sentence, and so on), emotion features (such as affect words, emotional tone, etc.), manipulation features (such as personal pronouns, impersonal, and so on).

Zubiaga et al. in [6] have explored content-based features such as word vectors, speech tags, and the ratio of message letters to total alphabetical tweets, the number of words, the use of question marks and the use of exclamation marks. Some of the machine learning algorithms used to implement the method are CRF (Conditional Random Classifier), which is a statistical classifier used in structured learning, logical regression classifier and query-based classifier. In their method, they used a dataset in twitter with some of the news features in the dataset. They have achieved a precision of 46% on PHEME dataset. Low accuracy of this method is due to the applying one simple classifier.

The authors in [16] have investigated a neural network to process the missing values and improve the data set presented in the context of fake news detection. The content-based features considered in this article are statement ID, subjects discussed by the speaker, title of the speaker's job and location of the speech. The positive point of this article is applying a neural network instead of conventional classifier that helps to reach the desired fake news detection accuracy. They improved the accuracy by more than 15%.

B. Fake news detection based on social context features

In this section, social context-based methods [19] have been studied briefly. These studies for detecting fake news utilized social features that refer to the features of the user profile and user behavior in social media [20].

Buntain et al. [21] proposed a method to classify popular twitter messages into fake and true news based on social context with structural, content and temporal features. Structural features are twitter specific features for tweets e.g., rate of retweets or media shares. Temporal features describe previous features over time e.g., average author age over time.

For the classification, they investigated 100-Tree Random Forests. They achieved an accuracy of 66.93% in PHEME dataset.

Yang et al. [22] explored an unsupervised method to find the fake news and user's credentials using an unlabeled dataset. They proposed a Collapsed Gibbs Sampling method. They utilized user engagement features (such as likes, retweets and replies). They achieved an accuracy of 75% in LIAR dataset and 67% in BuzzFeed dataset.

Jin et al. [23] investigated conflicting viewpoints in a credibility propagation network for verifying news. They applied an unsupervised topic method with status features (which indicates people's response to the message, such as message support, message rejection, etc.) to find out conflicting viewpoints. They analyzed experiments on a dataset collected from Sina Weibo. Kaliyar et al. [24] explored a novel approach that utilized social context features with content features. They proposed to combine different parallel blocks of single-layer deep Convolution Neural Network to detect fake news accurately. They accomplished high accuracy in FakeNewsNet dataset.

C. Deep neural network for fake news detection

Deep learning method is also an attractive method for classification applications. In recent years, several deep learning methods have been developed to detect fake news. These approaches detect fake news without feature engineering and for this reason, accuracy of approaches is improved [25].

Ruchansky et al. [26] combined three characteristics of fake news including parties, users and the article. In their paper, a hybrid method is introduced by combining three steps. In the first step, called capture, a Long Short-Term Memory (LSTM) algorithm determines the temporal characteristics pattern. The second step, the score, uses singular value decomposition to determine the behavioral characteristics of social media users. Finally, the results of the two methods are combined in a way that the resulting output is used for classification. They have experimented their approaches on dataset collected from Twitter and Sina Weibo.

Another deep learning method is weighted sum method [27]. In the paper, content features are extracted by a word embedding algorithm for timely fake news detection. The presented method for fake news detection works through a two-path CNN in a way that one path contains weighted sum of shared CNN and supervised CNN and the other path contains weighted sum of shared CNN and unsupervised CNN. They achieved a precision of 44.20% on PHEME dataset. The important point in this paper is that classifiers fused with the weighting method that increased the accuracy compared with voting fusion.

Mahabub et al. [28] proposed a hybrid method based on voting. In this study, several machine learning methods was implemented and their outputs were compared. Then, among them, the three algorithms that had the best outputs based on the accuracy metric were combined with the ensemble voting method. The three algorithms were selected as follow: Multilayer Perceptron, logistic regression and X-Gradient Boosting. Linguistic features were also used. This study performed on a dataset collected from BuzzFeed and PolitiFact. The drawback of the method was using voting fusion which reduced accuracy of detecting.

Kaliyar et al. [29] explored a novel approach. In this study, text features were converted to vectors with GloVe method and then these vectors were processed with a deep neural network approach. Their method had 3 convolutional layers with different kernel sizes (filter sizes) that helped to yield high accuracy in detecting fake news. Performance of this approach was evaluated on Kaggle fake news dataset.

Kaliyar et al. [30] discussed a deep neural network with five dense layers and different kernel sizes in each layer for detecting fake news. They utilized content, social context, and user-community-based features. They also achieved an accuracy of 92% in PolitiFact and 91% in BuzzFeed datasets.

Goldani et al. [31] investigated capsule neural network for prediction. Capsule neural network uses inverse engineering for classification and works better than Convolution Neural Network. They investigated text features on LIAR and ISOT datasets.

As described, the challenges that all the methods presented in this section are faced are selection of the features and classification methods that properly classify the fake news. In the method presented in the next section, we use context features and a fusion method to use not very complex classifiers for fake news classification. Then, for increasing the detection accuracy, OWA fusion method that was introduced by Ronald Yager [32] is used with a method for weight determination. Details are presented in the next section.

III. THE PROPOSED METHOD

As mentioned, it is not possible to determine whether a news event is fake or not confidently just

by using the news content features. Rather, to identify fake news, in addition to the message features, user profile features should also be considered. Therefore, detecting fake news needs identifying the users involved, extracting useful features from the messages and the user's profiles, and using network interactions.

To define the problem, some definitions are needed. Social interactions are represented as a set of multiple elements $\varepsilon = \{e_{it}\}$. This set indicates how the news is delivered via n different users $U = \{u_1, u_2, \ldots . u_n\}$ in the corresponding published text messages as $P = \{p_1, p_2, \ldots . p_n\}$ at time t. Each interaction $e_{it} = \{u_i, p_i, \mathrm{t}\}$ indicates that the user u_i has sent a message in p_i format at time t. When a message is not yet interactive, t = *null* and so u_i represents the author of the message. The p_a includes name, id, the number of followers, the number of friends, the number of lists, and some other features. The text message c_a contains id, text, the number of retweets, the number of likes and some other features.

As inputs to the fake news detection system, news interactions ε in social media among $'n'$ users for news $'a'$ are given. The task of detecting fake news is defined as predicting whether the news message $'a'$ is part of a fake news item. Equation (1) describes the problem.

$$\mathrm{F} : \mathrm{E} \rightarrow \{0,1\} \tag{1}$$

$$\mathrm{F}(a) = \begin{cases} 1 & if\ 'a'\ is\ part\ of\ a\ fake\ news \\ 0 & otherwise \end{cases}$$

In equation (1) F is a prediction function [3].

The fake news detection framework proposed in this manuscript known as fake news detection by ordered weighted average fusion (FNDOWAF) consists of three steps: feature selection, pre-processing and detection. In feature selection, some of the features are selected from the target dataset and then preprocessed (e.g., normalization and noise elimination). Finally, in the detection step, a data fusion system, consists of a number of detection components (DCs), determine the probability of a message being fake. The DCs are trained based on the training data prepared in the early steps of making the system. In this manuscript, the presented method detects fake news in Twitter using a combination of message features and users profile features.

As indicated in Fig. 1, inputs to the system are streams of different tweets produced for some events. The following steps convert these inputs to a dataset suitable for the remaining steps of the system operation. Initially, all the fake event tweets and the actual event tweets are collected for the next step as they are received in a stream form or as a batch file. Details of the other steps of the proposed method are as follow.

Message features indicate people's reaction to a message posted in social media. People's reaction to fake news is more intense than real news because the discussions about a fake news context are more controversial and argumentative [33]. Moreover, according to the analysis done in [34], verified social media users are more likely to spread real news. The users that send more fake news also have fewer followers and more friends. Because people who publish fake news need a larger network to publish more news, they ask for more friends. On the other hand, users who publish fake news send fewer messages and people who post real news are more active in social media.

In order to detect the fake news, in the feature selection step, a combination of user features and message features is considered, as illustrated in Fig. 2 and Fig. 3. Table 1 also describes these features with more details.

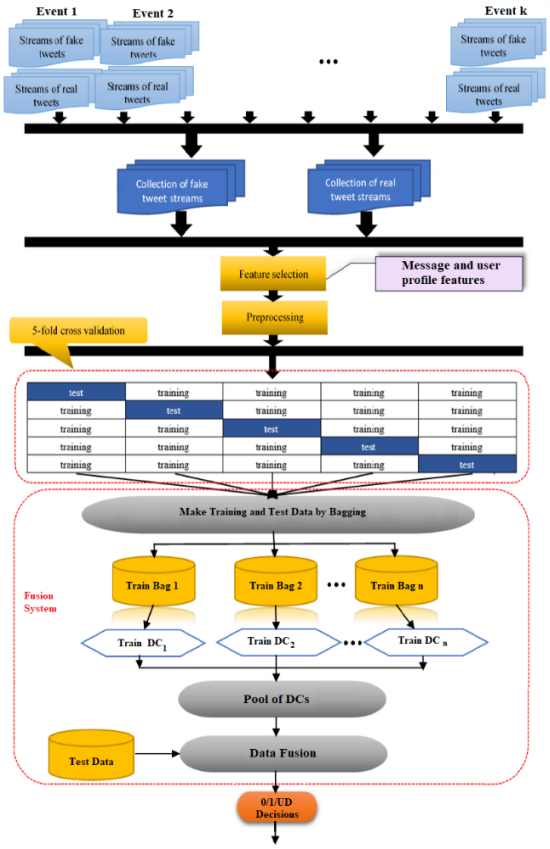

Fig. 1. Overview of the FNDOWAF method for the detection of fake news.

Fig. 2. A Twitter user profile and major features of the user.

Fig. 3. A tweet and major message features.

TABLE I. DESCRIPTION OF SOME OF THE SOCIAL FEATURES

Feature	Description
The number of friends	The number of users that the twitter user has sent a friend request to them.
The number of followers	The number of users who have sent a friend request for the user on Twitter.
The number of lists	The number of groups and lists in which the user is a member.
The number of statuses	The number of messages a user has posted on Twitter since joining.
Verified	Whether or not a user is verified of Twitter, which appears as a blue mark in the corner of the user's profile.
The number of retweets	The number of times the message is sent by other users on Twitter.
The number of favorites	The number of times the message is liked by Twitter users.

In the preprocessing step, after receiving all the tweets as a bulk of records, the rows that contain missing values are deleted, and the data are normalized. The normalization operation maps all the values of all the features into $[0,1]$ interval. This normalization increases the performance of training the base classifiers known as Decision Components (DCs) in the fusion system.

In the fusion step, a number of DCs are used. Each of the DCs estimate the probability of an incoming tweet being fake. There are some points about the DCs that should be considered. The number of DCs is selected heuristically such that a tradeoff between complexity and accuracy is obtained [35]. Using more DCs increases the accuracy and reliability of estimation while fusion complexity is increased. On the other hand, the structure and behavior of the DCs should be different because using any number of the same DCs does not improve the accuracy of the fusion system. Therefore, using diverse DCs is a vital property of the fusion system [36].

Diversity in DCs is created in different ways such as using the DCs with different architectures, structures, and training the DCs with different training data [37]. In the proposed method, the DCs are neural networks with different number of layers, different number of neurons in each layer, different activation functions and the bagging [38] is used for training. Therefore, the diversity is guaranteed for the fusion system. Details of the fusion system are described next.

The base classifiers in the fusion system (the DCs in Fig. 1) are multi-layer perceptron with one or two hidden layers. The number of neurons in the input layer equals to the number of features selected from the input dataset as determined in the experiments section for each dataset. One output neuron produces an output in the range $[0,1]$ to determine the probability of the input message being fake. For training and testing each DC, 70% and 30% of the input tweets are selected respectively. The training algorithm uses bagging method to select training and validation data for each DC to keep them diverse. After training, the DCs are used in the fusion system as indicated in Fig. 1.

In the fusion step, OWA fusion method is used. The OWA method is introduced first by Ronald Yager [32]. Using OWA as a data fusion method has a few steps. Input to the fusion system is defined as a vector $X = (x_1 . x_2 . \dots . x_n)$ such that x_i is the output estimate produced by the i^{th} DC. Corresponding to the vector X, another vector $B = (b_1 . b_2 . \dots . b_n)$ is defined such that b_i is the i^{th} largest element in X. A weight vector is also considered as $W = (w_1 . w_2 . \dots . w_n)$ such that Equation (2) holds:

$$\sum_{i=1}^{n} w_i = 1 \ \ (0 \le w_i \le 1 \, . \, i = 1.2. \dots . n) \qquad (2)$$

The OWA operator with the weight vector W is defined as a mapping function F such that $F: R^n \to R$ where F is defined by equation (3).

$$F_w(B) = \sum_{i=1}^{n} w_i b_i \qquad (3)$$

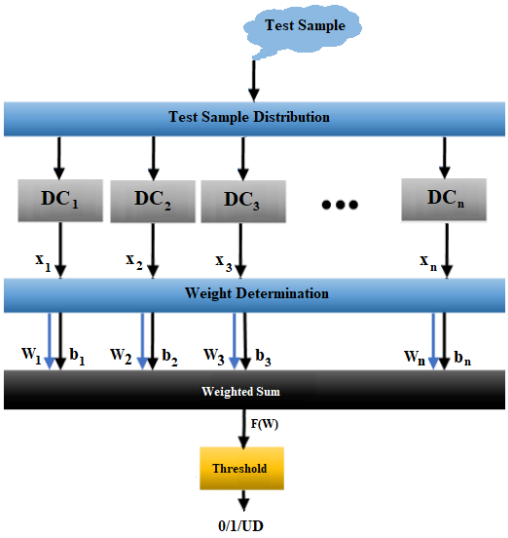

Fig. 4. The proposed Data Fusion system architecture

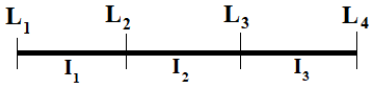

Fig. 5. Output intervals defined for each DC.

The main challenge of the OWA operator is determining the weight vector [39]. However, different methods have been introduced to compute the weights [40] [41]. The way that is introduced in this paper to compute the weights is described next. Fig. 4 indicates general structure of the fusion method used in this manuscript.

In Fig. 4, the output of each DC is $x_i \in [0,1]$, the output of the fusion system is the value of mapping function $F \in [0,1]$ and final decision is made by using a threshold function $\tau \in \{0,1,UD\}$. If $\tau = 0$ then the input sample is not a fake news while $\tau = 1$ indicates a fake news. If $\tau = UD$ 1then, no decision is made about the input sample, then, it is a suspicious one and needs more processing. The threshold function is defined later by details.

In the weight computation section, the input to the fusion system (X) is divided into three intervals as indicated in Fig. 5. The first interval I_1 includes all the values produced by the DCs that are near 0, the third interval I_3 includes all the values near 1, and the second interval I_2 includes all the middle values.

As indicated in Fig. 5, the output values of each DC are divided into three sections to be used for the computation of W. Based on the structure of the fusion system indicated in Fig. 4, n is the number of DCs, n_1 is the number of output values such that $DC_i \in [L_1, L_2]$, n_2 is the number of output values such that $DC_i \in [L_2, L_3]$ and n_3 is the number of output values such that $DC_i \in [L_3, L_4]$. For each test sample, all the outputs of the DCs are computed as vector X. Then, equation (4) is used to compute the weights in W.

$$v_i = \begin{cases} \frac{b_i - L_1}{L_2 - L_1} \times \frac{n_1}{n} & b_i \in [L_1.L_2] \\ \frac{b_i - L_2}{L_3 - L_2} \times \frac{n_2}{n} & b_i \in [L_2.L_3] \\ \frac{L_4 - b_i}{L_4 - L_3} \times \frac{n_3}{n} & b_i \in [L_3.L_4] \end{cases} \quad (4)$$

Based on equation (4), the components of W are computed by normalizing the values of v_i by equation (5).

$$w_i = \frac{v_i}{\sum_{j=1}^{n} v_j} \quad (5)$$

In the next section, some experiments are designed to compare the effectiveness of the proposed method with respect to some other ones. One of the fusion methods used for comparison is majority vote (MV). Fig. 6 indicates the structure of the MV fusion system by using the DCs.

As indicated in Fig. 6, each DC estimates the probability of the input tweet being spam. Then, by using a threshold value, the final decision is made (0 for not spam, 1 for spam, and UD for suspicious). Finally, a majority vote computes the final decision. In the next section, some experiments are designed to indicate the effectiveness of the proposed method as compared with some others. Details of the datasets used for the experiments and analysis of the results are also presented in the next section.

To summarize all the steps in the FNDOWAF, the pseudocode in Fig. 7 describes the details.

[1] Un Decided

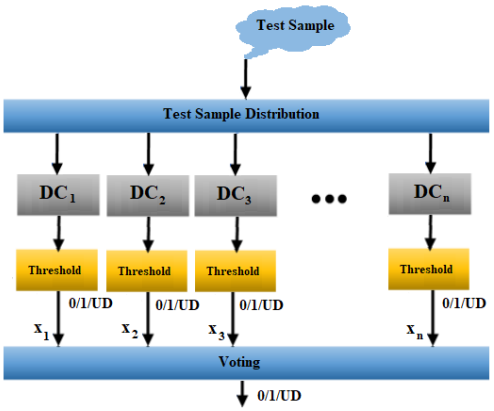

Fig. 6. MV fusion system architecture.

Algorithm: FNDOWAF

1- Load fake news dataset and real news dataset
2- Combine the two datasets and make a complete dataset
3- Select the most valuable profile and message features and make the message dataset for train and test.
4- In preprocessing stage, remove the rows With missing values, normalize each Feature and bring the values to the range [0,1].
5- Select 5-fold cross validation. Select 70% Of the data randomly for training the DCs In each fold and 30% for test.
6- Select the structure of each DC by selecting The number of hidden layers and the number Of neurons in each layer randomly. Parameters Of the activation functions for these neurons Are also selected randomly.
7- For training each DC, bagging is used to select Training data for each DC and the training is Done.
8- For the test samples, OWA weights are computed
9- By equation 3 and the output is produced in the Fusion system as indicated in Fig. 7 and error Is computed.
10- Repeat steps 5 to 10 for all the folds.
11- Average the results obtained in the folds to Compute final results.

Fig. 7. pseudocode for the FNDOWAF.

II. EXPERIMENTS AND ANALYSIS

In this section, the dataset used for the experiments is introduced. Then, the results of the Experiments are analyzed. In the experiments in this section, the PHEME [6] dataset is used. To prepare the dataset for the experiments, reputable individuals and journalists capable of detecting fake news are selected first, and the news are collected in a variety of ways. According to the journalist comments, this news is divided into two categories: fake and real. In the selected dataset, eight most viewed events are selected as follow:

• **Charlie Hebdo shooting:** A terrorist attack took place on January 1, 2008, by two gunmen at the Charlie Hebdo comic book office. Two people were killed and four others were wounded in the shooting.

• **Ferguson unrest:** August 9, 2014 Citizens of Ferguson, Michigan, in the United States, demonstrated after a deadly shooting of a white police officer by an 18-year-old black boy.

• **German wings plane crash:** A passenger plane traveling from Barcelona to Dusseldorf crashed in the French Alps on March 24, 2015, killing all passengers and guests. The plane was deliberately crashed by one of the pilots.

• **Gurlitt collection:** In November 2014, there was a rumor that the Bern Museum of Fine Arts was going to buy a collection of masterpieces from the son of a Nazi Germany dealer. Eventually, the museum confirmed the rumor by purchasing a collection of artifacts.

• **Ottawa shooting:** On October 22, 2014, a shooting at a Canadian parliament in Ottawa killed a Canadian soldier.

• **Prince to play in Toronto:** On November 3, 2014, a rumor was circulated that Prince (the singer) played a secret show in Toronto that night. Some people even attended the concert, but the rumor was later confirmed.

• **Sydney siege:** On December 15, 2014, a gunman kidnapped 10 customers and 8 employees by attacking a Lindt chocolate cafe located at Martin Place in Sydney, Australia, in December 15, 2014.

• **Putin missing:** In March 2015, rumors circulated about the 10-day absence of Russian President Vladimir Putin that on the eleventh day, Putin ended all rumors about his death and illness in public.

The number of fake and real tweets in each event is shown in Table 2. As shown in the table, there are 2458 fake news and 4023 real news in this dataset. By collecting the news in a dataset, there are 6481 news at all. By removing noise and outliers from this data set, there are finally 6340 news with a tag field as 1 or 0 for fake or real respectively.

In the experiments, the value of two parameters influences the results of the experiments. These two parameters that are a threshold value and the number of input features to the experiments, should be determined.

Basically, the threshold is used to convert the probability value produced by the DCs, which is in the interval $[0,1]$, to a decision as $\{0,1,UD\}$ to determine that the sample tweet is not spam, is spam, or is undecided respectively. This is because the PHEME dataset contains two categories of news that are labeled as fake or not fake. Therefore, to compare the output of the proposed method with the actual value in the dataset, the threshold is used in the output of the system.

TABLE II. THE NUMBER OF FAKE AND REAL NEWS FOR 8 NEWSWORTHY EVENTS

Event	Fake news	Real news	total
Charlie Hebdo shooting	458 (22%)	1621 (78%)	2079
Sydney siege	522 (42.8%)	699 (57.2%)	1221
Ferguson unrest	284 (24.8%)	859 (75.2%)	1143
Ottawa shooting	470 (52.8%)	420 (47.2%)	890
German wings plane crash	238 (50.7%)	231 (49.3%)	469
Prince to play in Toronto	299 (98.7%)	4 (1.3%)	303
Putin missing	126 (53%)	112 (47%)	238
Gurlitt collection	61 (44.2%)	77 (55.8%)	138
total	**2458 (38%)**	**4023 (62%)**	6481

In the designed experiments, the structure of the DCs is neural network, each one with different number of layers, neurons, activation functions and different sets of training data to make them diverse for the fusion system as described in the last section. The dataset is divided into train, test and validation parts with 50%, 20% and 30% of the samples respectively. The train and test parts are used to train and test the DCs. The validation part is used to validate the operation of entire fusion system. For diversity reason, bagging is used for training the DCs.

In the next experiments, 30 DCs were trained and used by the fusion system. For the experiments, not only the number of layers and neurons in each layer, but also the type of activation functions is selected randomly. It should also be noted that the effect of the number of DCs in the accuracy of the system is also investigated later in this section. In this experiment, the output values produced by the DCs for the sample tweets, are analyzed to determine the thresholds that separate the fake news from the others. The results of the first experiment are shown in Fig. 8.

As illustrated in Fig. 8, most of the values produced by the system for the fake news are in the interval $[0.38 \dots 0.55]$ while for the real news, these values are in the interval $[0.35 \dots 0.45]$. By comparing these two intervals, it is possible to exclude from the interval $[0.35 \dots 0.55]$ the common values for fake news and real news. Therefore, different interpretations for the intervals indicated in Fig. 9 is possible to determine fake (1) and not fake (0) news as follow:

A. $[0.38 \dots 0.45]$: values less than 0.38 are mapped into 0 and greater than 0.45 into 1
B. $[0.39 \dots 0.44]$: values less than 0.39 are mapped into 0 and greater than 0.44 into 1
C. $[0.40 \dots 0.43]$: values less than 0.4 are mapped into 0 and greater than 0.43 into 1
D. $[0.41 \dots 0.42]$: values less than 0.41 are mapped into 0 and greater than 0.42 into 1
E. $[0.415]$: the values less than or equal to 0.415, are mapped into 0 otherwise into 1.

It should be noted that for the output values that fall inside each of the intervals, no decision is made and the input sample is marked as undecided (UD). The main difference in the intervals A-E is the reliability of the decisions. If the interval A is used for decision making, which is the widest one, the samples with the highest reliability are marked as 0 or 1 and the reliability of the decision is also the highest. In contrast, if the interval E is used, sharp decisions are made and no samples are marked as UD, therefore, least reliability is obtained. In decision making, a moderate interval may be used to produce reliable decisions while least number of samples are marked as UD.

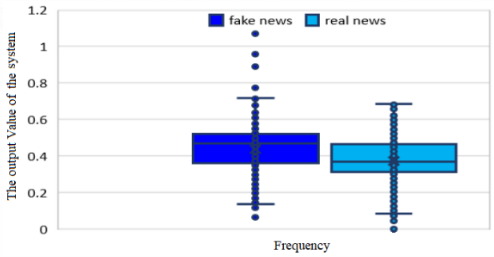

Fig. 8. Distribution of output values of the system for fake and real New.

In the next experiments, three different metrics are used to compare the results of the methods used in the experiments (Precision, Recall, and F1_score). To compute the values of these metrics, four parameters are defined. TP[2] indicates the number of test samples that are fake news and correctly classified as fake. TN[3] indicates the number of test samples that are not fake news and correctly classified as not fake. FP[4] indicates the number of test samples that are not fake news but incorrectly classified as fake. FN[5] indicates the number of test samples that are fake news but incorrectly classified as not fake. Equations (6), (7) and (8) compute the values of the three metrics.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{F1\_Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

Fig. 9 indicates the results of the experiments by using intervals A-E and the values of the three metrics.

As Fig. 9 indicates, the proposed method is compared with the CRF method [6] by using the intervals A-E. In this Fig., by using each of the intervals, the value of precision, recall, and F1-score is more than the CRF method. The difference among the cases that the intervals A-E are used is the number of samples that are marked as UD. For the sharp interval E, no sample is marked as UD and for all the samples, a 0/1 decision is made while in contrast, for the widest interval A, maximum number of samples are marked as UD because maximum widths interval is used. However, in all the cases, the FNDOWAF has higher performance than the CRF method.

Fig. 10 indicates the number of samples that are marked as UD by selecting any one of the intervals A-E.

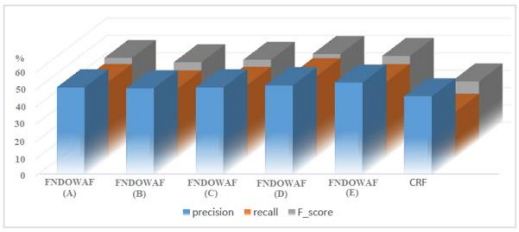

Fig. 9. Precision, Recall and F1_Score with different thresholds.

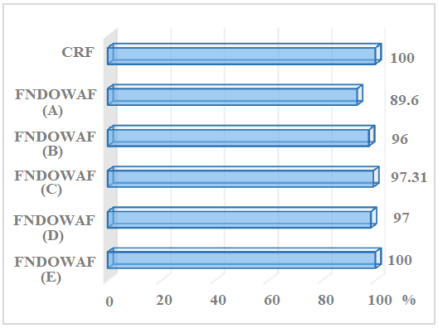

Fig. 10. Percent of the data marked as UD by using different thresholds.

As indicated in Fig. 10, by using the sharp interval (E), no samples are marked as UD. On the other hand, by using interval A, which is the widest one, 10.4% of the samples are marked as UD.

In the next experiment, the effect of deleting each feature from the input sample in the performance of the FNDOWAF is investigated. The metric that is used in this experiment is mean square error (MSE) as computed by equation (9).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(d_i - l_i)^2 \tag{9}$$

In equation (9), n is the number of test samples, d_i is the output value produced by the fusion system and l_i is the actual label of the i^{th} sample (0/1). The results of the experiment are shown in Fig. 11.

[2] True Positive

[3] True Negative

[4] False Positive

[5] False Negative

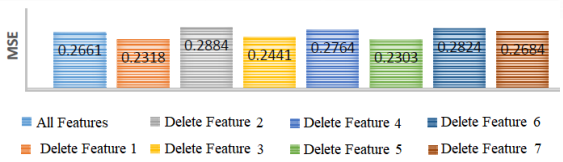

Fig. 11. Feature elimination with mean square error metric.

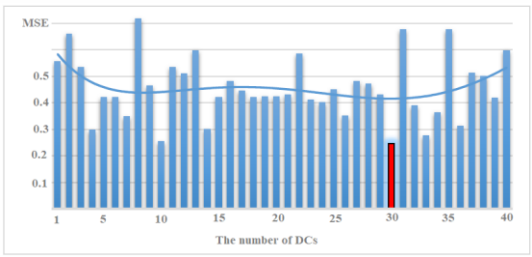

Fig. 12. The effect of the number of DCs on MSE.

Evaluating the results in Fig. 11 reveals the fact that some of the features of the samples dataset have positive effect (reducing the value of error) while some others have negative effect (increasing the value of error) on the output results of the FNDOWAF method. The seven features in Fig. 11 are numbered 1-7 and labeled as the number of likes, the number of retweets, the number of followers, the number of lists, status, the number of friends and verification. In Fig. 11, removing feature 2 (the number of retweets) makes maximum increase in the error. Therefore, this feature has the most positive effect in error reduction. In contrast, removing feature 5 decreases the error more than others, therefore, this feature does not improve the effectiveness of the method.

In any DF system, the number of components or decision makers also play an important role in the DF performance by the time the diversity is considered. To indicate the behavior of the system when different number of DCs are used, the next experiment is designed. Fig. 12 indicates the effect of increasing the number of DCs on the total error of the system.

In Fig. 12, the DCs are added one by one from 1 to 40 to evaluate the effect of this number on the performance of the proposed method. In this experiment, in each step, a DC is selected and the number of layers and the number of neurons is randomly selected with a random activation function. Then, the DC is trained with the training data selected by the bagging method to guarantee the diversity. Moreover, the DCs are trained to have moderate, not minimum error to have more diversity. In Fig. 12, it is expected that by increasing the number of DCs, the error is decreased. But, because of the randomness in design and selection of DCs structures and training, in some cases, the error may increase. Moreover, as the number of DCs is increased, it is also expected that the behavior of the fusion method being smoother.

In the next experiment, the FNDOWAF method is compared with some other fusion methods. Except CRF and DTSL, the other methods are implemented and tested with the data used for FNDOWAF. Fig. 13 uses precision as a metric for comparison as computed by equation 6. The interval that is used in this experiment is D which is a moderate one.

In Fig. 13, when no fusion is used, only the output of a single DC is used for all the validation samples and the precision is computed. As indicated by Fig. 13, the proposed method has maximum precision among the selected methods. Fig. 14 compares the methods by using recall metric computed by equation (7).

As indicated in Fig. 14, the value of recall for the proposed method is near the voting and more than the others. Since in the proposed method, some of the test samples are labeled as UD because the interval D is used, a small decrease in recall is obtained with respect to the voting method. In Fig. 15, F1_score is compared for the selected methods.

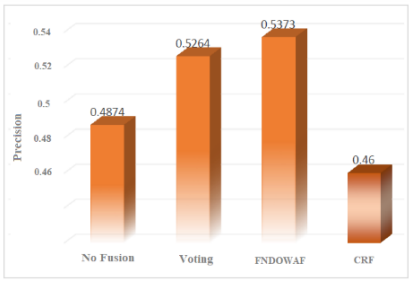

Fig. 13. Comparing precision of FNDOWAF with other methods.

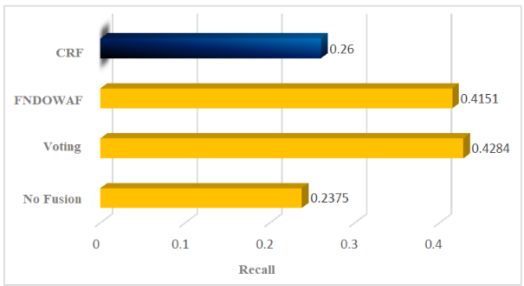

Fig. 14. Comparing recall metric.

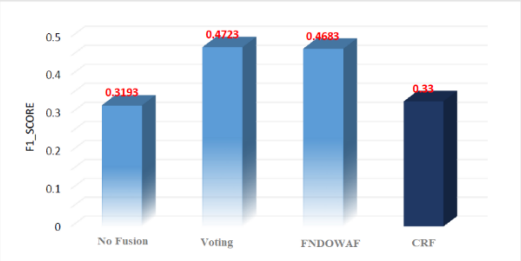

Fig. 15. Comparing F1_score metric

As indicated in Fig. 15, the value of the proposed method is near the voting and is near the maximum among the selected methods. Because in the proposed method, some of the samples are labeled as UD, a small decrease in the value of F1_score is seen. These samples are suspicious and it is not possible to decide about fakeness of them with high confidence. If more processing is done on such samples, the value of the metrics shown in Fig. 13, 14, 15 also have more increases for the FNDOWAF. In general, the FNDOWAF method improved the precision, recall and F1_Score by 7%, 16% and 13%, respectively, compared to the CRF method.

As a conclusion, the last experiment, compares the previous results with a deep learning method ran on the same dataset. Table 3 indicates the results of the three metrics used for comparison. As shown in this table, the precision of the proposed ensemble method is higher than the others.

A. Discussion

The experiments in this section are designed in two categories. In the first category, there are some experiments that indicate how different parameters of the proposed method affect the performance. In the second category, the experiments compare the results of the proposed method with others.

In the first category of the experiments, the main parameters of the method that are examined are the output threshold, the number of DCs, and the number of features used in the fusion system. As indicated in Fig. 9, the effect of the threshold used in the output of the fusion system to make a 0/1 decision is more on recall than precision and F-score.

The effects of different features on the output results are indicated in Fig. 11. As indicated, some features have more effects on the results such that by eliminating them, an increase in the MSE occurs. The effect of the number of DCs is indicated in Fig. 12. As a general rule, the more the number of DCs, the lower the value of error. This decrease has a limit, after that, the error begins to increase again. Therefore, a moderate value for the number of DCs should be selected. Moreover, as the number of DCs is increased, due to the randomness of the structure and training of the DCs, in some cases, the error may increase.

In the second category of experiments, the method is compared with some other known methods. Figs. 13, 14, 15 compare the precision, recall and F1-score obtained by the proposed method with the others. As seen, these are improved by the proposed method.

In table 3, the precision of the proposed method is higher than all the others. Although the two other metrics are higher for the DTSL method, it should be noted that the deep learning methods are very complex and need large amounts of training data while, their time complexity is also very high. In contrast, the presented method uses simple base classifiers and fusion method with lower time complexity and lower number of training samples. Another reason that may decrease the values of recall and f-score is that in the proposed method, some of the samples may be marked as undecided that need more processing.

Most studies in the field of fake news detection have primarily considered content features. If the fake news was written with the intent to sabotage, the content features alone may not be able to accurately detect the fake news. In addition, the number of content features are high and slows down the processing. Therefore, in our proposed method, an attempt has been made to detect fake news with a smaller number of social context features. These features indicate the relationship between news and user features on social media.

TABLE III. COMPARISON WITH EXISTING BENCHMARKS WITH PHEME

Authors	Precision (%)	Recall (%)	F-score (%)
Dong et al. [27] (DTSL)	44.20%	77.58%	57.98%
Zubiaga et al. [6] (CRF)	46.2%	26.8%	33.9%
Proposed model-No Fusion	48.74%	23.75%	31.93%
Proposed model-Voting	52.64%	42.84%	47.23%
Proposed model-FNDOWAF	53.73%	41.51%	46.83%

IV. CONCLUSIONS

In this manuscript, an effective method is introduced to detect fake news in social media. The news published in social media has features including news content features, linguistic features, visual features, and social features. In the proposed method in this manuscript, the social features of the news, which include message features and user profile features are used to detect the fake news.

The proposed method attempts to detect fake news by using a fusion method. By using a dataset from the twitter, for each input sample, which contains the features extracted from the tweets and profiles of the users, the proposed fusion system produces a three valued decision with values 0/1/UD corresponding to normal, fake or undecided. If the output value is 0, then the tweet is not fake while if it is 1, the tweet is labeled as fake and if it is UD, it means that it labeled as undecided, therefore, more processing is needed for a final decision making.

Different experiments are designed to indicate effectiveness of the method with respect to the selected methods. The metrics that are used for comparison are precision, recall, f1_scor, and mean square error. The results indicate an increase in performance by comparing these metrics. Our future research will concentrate on using more features and introducing new fusion methods to increase the accuracy of the system.

REFERENCES

[1] M. Salkhordeh Haghighi and E. Mohammad, "Introducing a genetic algorithm based Method for Community person's stance Detection in social media and news," *Journal of Information Systems and Telecommunications (JIST)*, vol. 11, no. 41, p. 18, 2020.

[2] H. S. Cheraghchi and A. Zakerolhoseini, "COGNISON: A Novel Dynamic Community Detection Algorithm in Social Network," *Journal of Information Systems and Telecommunications (JIST)*, vol. 4, no. 2, 2016.

[3] A. S. S. W. J. T. a. H. L. Kai Shuy, "Fake News Detection on Social Media:A Data Mining Perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22-36, 2017.

[4] H. allcott and M. gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives—*, vol. 31, no. 2, pp. 211-236, 2017.

[5] Y. C. N. J. C. Victoria L. Rubin, "Deception detection for news: three types of fakes," in *2015 Annual Meeting of the Association for Information Science and Technology*, USA, 2015.

[6] M. L. R. P. Arkaitz Zubiaga, "Learning reporting dynamics during breaking news for rumour detection in social media," University of Warwick, Coventry, UK, 2016.

[7] M. L. R. P. G. W. S. H. P. T. Arkaitz Zubiaga, "Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads," University of Warwick, UK, 2016.

[8] V. L. R. Y. C. Niall J. Conroy, "Automatic Deception Detection: Methods for Finding Fake News," in *ASIS&T2015*, St. Louis, MO, USA, 2015.

[9] H. Z. W. P. Zhongyue Zhou, "weibo rumor detection method based on user and content relationship," in *Artificial Intelligence in China: Proceedings of the International Conference on Artificial Intelligence in China*, china, 2020.

[10] J. V. d. S. J. G. J. F. M. d. S. F. A. M. d. O. J. J. F. d. Souza, "A systematic mapping on automatic classification of fake news in social media," *Social Network Analysis and Mining*, 2020.

[11] O. A. A. O. Bodunde Akinyemi, "An Improved Classification Model for Fake News Detection in Social Media," *I.J. Information Technology and Computer Science, MECS*, vol. 1, pp. 34-43, 2020.

[12] N. R. M. G. A. B. Harita Reddy, "Text-mining-based Fake News Detection Using Ensemble Methods," *International Journal of Automation and Computing (2020)*, 2020.

[13] J. J. L. M. S. K. H. L. C. a. R. X. Jianfei Yu, "Coupled Hierarchical Transformer for Stance-Aware Rumor Verification in Social Media Conversations," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 2020.

[14] Q. Z. L. S. Quanzhi Li, "Rumor Detection By Exploiting User Credibility Information, Attention and Multi-task Learning," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019.

[15] A. A. Anshika Choudhary, "Linguistic Feature Based Learning Model for Fake News Detection and Classification," *Elsevier Ltd*, 2020.

[16] X. D. N. L. L. Q. Chandra Mouli Madhav Kotteti, "Fake News Detection Enhancement with Data Imputation," in *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress*, Athens, Greece, 2018.

[17] J. K. K. R. J. B. B. S. Martin Potthast, "A Stylometric Inquiry into Hyperpartisan and Fake News," *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, vol. 1, p. 231–240, 2018.

[18] I. G. D. S. Vivek K. Singh, "Detecting fake news stories via multimodal analysis," *Association for Information Science and Technology*, pp. 1-15, 2020.

[19] C. K. M. J. G. X. Z. a. R. Z. Niraj Sitaula, "Credibility-Based Fake News Detection," in *Disinformation, Misinformation, and Fake News in Social Media*, Springer, Cham, 2020, pp. 163-182.

[20] J. N. J. T. Andrew Guess, "Less than you think: Prevalence and predictors of fake news dissemination on Facebook," *American Association for the Advancement of Science*, vol. 5, 2019.

[21] J. G. Cody Buntain, "Automatically Identifying Fake News in Popular Twitter Threads," in *2017 IEEE International Conference on Smart Cloud*, New York, NY, USA, 2017.

[22] K. S. S. W. R. G. F. W. H. L. Shuo Yang, "Unsupervised Fake News Detection on Social Media: A Generative Approach," in *33rd AAAI Conference on Artificial Intelligence*, Hilton Hawaiian Village, Honolulu, Hawaii, USA, 2019.

[23] J. C. Y. Z. J. L. Zhiwei Jin, "News verification by exploiting conflicting social viewpoints in microblogs," in *AAAI'16 Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, Arizona, 2016.

[24] A. G. &. P. N. Rohit Kumar Kaliyar, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimedia Tools and Applications volume 80*, p. 11765–11788, 2021.

[25] E. A. M. G. Sherry Girgis, "Deep Learning Algorithms for Detecting Fake News in Online Text," in *13th IEEE International Conference on Computer Engineering and Systems*, Cairo, Egypt, 2018.

[26] S. S. Y. L. Nataly Ruchansky, "CSI: A Hybrid Deep Model for Fake News Detection," in *CIKM '17: Proceedings of the 2017 ACM on Conference on*

Information and Knowledge Management, Singapore, 2017.

[27] U. V. S. C. L. Q. Xishuang Dong, *Deep Two-path Semi-supervised Learning for Fake News Detection*, USA: arXiv.org, 2019.

[28] A. Mahabub, "A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers," *Springer Nature Switzerland AG 2020, SN Applied Sciences*, 2020.

[29] A. G. P. N. S. S. Rohit Kumar Kaliyar, "FNDNet – A deep convolutional neural network for fake news detection," *Cognitive Systems Research*, vol. 61, pp. 32-44, 2020.

[30] A. G. &. P. N. Rohit Kumar Kaliyar, "EchoFakeD: improving fake news detection in social media with an efficient deep neural network," *Neural Computing and Applications*, 2021.

[31] S. M. R. S. Mohammad Hadi Goldani, "Detecting fake news with capsule neural networks," *Applied Soft Computing*, vol. 101, 2021.

[32] R. Yager, "On Ordered Weighted Averaging Operators in Mulicriteria Decisionmaking," *Ieee Transactions on Systems, Man and Cybernetics*, vol. 18, no. 1, 1988.

[33] W. H. C. X. L. C. J. Z. Gang Liang, "Rumor Identification in Microblogging Systems Based on Users' Behavior," *IEEE Transactions on Computational Social Systems*, vol. 2, no. 3, pp. 99 - 108, 2015.

[34] K. Shu, S. Wang and H. Liu, "Understanding User Profiles on Social Media for Fake News Detection," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, Miami, FL, USA, 2018.

[35] M. Salkhordeh Haghighi and S. Kahrobaee, "Applying a fuzzy interval ordered weighted averaging aggregation fusion to nondestructive determination of retained austenite phase in D2 tool steel," *NDT and E International*, vol. 103, pp. 39-47, 2019.

[36] M. Salkhordeh Haghighi, A. Vahedian and H. Sadoghi Yazdi, "Creating and measuring diversity in multiple classifier systems using support vector data description," *Applied Soft Computing*, vol. 11, no. 8, pp. 4931-4942, 2011.

[37] I. Ahadi Akhlaghi, M. Salkhordeh Haghighi, S. Kahrobaee and M. Hojati, "Prediction of chemical composition and mechanical properties in powder metallurgical steels using multi-electromagnetic nondestructive methods and a data fusion system," *Journal of Magnetism and Magnetic Materials*, vol. 498, p. 166246, 2020.

[38] L. I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, Newjersey: John Wiley & Sons, 2014, p. 384.

[39] S. L. Y. Y. Lifeng Wu, "A Model to Determine OWA Weights and Its Application in Energy Technology Evaluation," *International Journal of Intelligent Systems*, vol. 30, p. 798–806, 2015.

[40] S. Vluymans, N. M. P. Mac Parthaláin, . C. Cornelis and Y. Saeys, "Weight selection strategies for ordered weighted average based fuzzy rough sets," *Information Sciences*, vol. 501, pp. 155-171, 2019.

[41] Z. Xu, "An overview of methods for determining OWA weights," *International journal of intelligent systems*, vol. 20, 2005.

Mehdi Salkhordeh Haghighi received his B.Sc. degree in 1994, M.Sc. in 1997 and Ph.D. in 2012, all in Computer Software Engineering in Ferdowsi University of Mashhad in Iran. He has been a member of teaching staff in Faculty of Computer Engineering of Sadjad University since 2002. He also has been a member of IEEE since 1990. His research interests are Data Fusion, Situation Awareness, Data Mining and Social Networks.

Nasim Eshaghian received her B.Sc. degree in Information Technology Engineering from Shahrood University of Technology, Semnan, Iran, in 2014. She also received her M.Sc. degree in Security Engineering from Sadjad University in Mashhad, Iran, in 2019. Her current research interests are Big Data Mining, Social-Media and Machine Learning.