IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 9, 2020, accepted August 21, 2020, date of publication August 26, 2020, date of current version September 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3019600

Lies Kill, Facts Save: Detecting COVID-19 Misinformation in Twitter

MABROOK S. AL-RAKHAMI[1,2], (Member, IEEE), AND ATIF M. AL-AMRI[1,3], (Member, IEEE)

[1]Research Chair of Pervasive and Mobile Computing, King Saud University, Riyadh 11543, Saudi Arabia
[2]Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[3]Software Engineering Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding author: Mabrook S. Al-Rakhami (malrakhami@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research, Research Chair of Pervasive and Mobile Computing, King Saud University, Riyadh, Saudi Arabia.

ABSTRACT Online social networks (ONSs) such as Twitter have grown to be very useful tools for the dissemination of information. However, they have also become a fertile ground for the spread of false information, particularly regarding the ongoing coronavirus disease 2019 (COVID-19) pandemic. Best described as an infodemic, there is a great need, now more than ever, for scientific fact-checking and misinformation detection regarding the dangers posed by these tools with regards to COVID-19. In this article, we analyze the credibility of information shared on Twitter pertaining the COVID-19 pandemic. For our analysis, we propose an ensemble-learning-based framework for verifying the credibility of a vast number of tweets. In particular, we carry out analyses of a large dataset of tweets conveying information regarding COVID-19. In our approach, we classify the information into two categories: credible or non-credible. Our classifications of tweet credibility are based on various features, including tweet- and user-level features. We conduct multiple experiments on the collected and labeled dataset. The results obtained with the proposed framework reveal high accuracy in detecting credible and non-credible tweets containing COVID-19 information.

INDEX TERMS Classification, COVID-19, machine learning, misinformation, Twitter.

I. INTRODUCTION

In recent months, the spreading rate of coronavirus disease 2019 (COVID-19) has been very high around the world [1]–[3] For most people, it is an unprecedented global pandemic in present times. These sentiments have been echoed by the World Health Organization (WHO) Director-General, Tedros Adhanom Ghebreyesus [4]. During the Munich Security Council held on February 15, 2020, he stated that the world was in a war to fight not only a pandemic, but also an infodemic. Various sources have shown that the spread of COVID-19 has been accelerated by the kinds of information, or indeed misinformation, being communicated [3], [4]. This is because individuals act to protect themselves from contracting the disease based on the information they receive from different sources. An example that illustrates the extent to which false information regarding the pandemic has spread was provided in a report by the International Fact-Checking Network (IFCN). ICFN brings together more than a hundred organizations that essentially do fact-checking. By April 2020, they revealed that over 4000 false claims had been spread regarding the pandemic [7]. Unfortunately, misinformation has a horde of adverse outcomes associated with it: it can increase fear of the pandemic among people [8], and it can mislead people when it comes to correct medical practices, as some may follow wrong advice on how to protect themselves from COVID-19, which could lead to ill health or even death [9], [10].

One of the ways through which misinformation is spread is through online social networks (OSNs) such as Twitter [11]. The spread of false information has been seen for other recent epidemics, such a Ebola [12], yellow fever [13], and Zika [14]. This is undoubtedly an alarming trend, since misinformation may be accepted by many individuals as correct information. To curb this ''misinformation challenge'', the WHO established the Mythbusters platform,

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

FIGURE 1. Two example tweets about COVID-19. (a) The left tweet contains misinformation regarding a cure for COVID-19, while (b) the right tweet represents credible information indicating that garlic has not been proved to be a cure against COVID-19.

which attempts to demystify false information regarding the COVID-19 pandemic.

These efforts have been complemented by individual fact-checking organizations that have come together under IFCN to increase the efforts of fighting the current infodemic [15].

OSNs have always been associated with the spread of misinformation, where many opinions appear to disregard scientific evidence and material in their posts [16]. A fundamental example that illustrates this is the myriad conversations regarding vaccines and cures to COVID-19. Many people have spread false information regarding COVID-19 cures, and have shown inclination towards accepting it. Figure 1 shows example tweets that promotes false COVID-19 cure and encourages people to purchase it. Another blatant example was the tweet by US President Donald J. Trump that suggested that combining hydroxychloroquine and azithromycin was the ''biggest game-changers in the history of medicine'' [17]. Medical bodies and medical research institutions subsequently came forth to demystify this false assertion, asking physicians not to follow such information when treating patients. Intended or otherwise, misinformation brings about fear and wrong medical prescriptions. It also leads to defiance of medical guidelines regarding preventive measures, such as social distancing and personal hygiene. This is as the direct result of a horde of misinformation [18].

As a very popular OSN platform, Twitter has become a tool for the spread of misinformation regarding COVID-19 [19], which has resulted in increased mental distress, anxiety, and self-harm. A worrying factor is that the rate at which misinformation is spreading surpasses physical distances, so that the fear of the pandemic spreads alarmingly fast, and seems to be more contagious than the actual COVID-19 [20]. As a result, patients and non-patients are affected both mentally and physically. Therefore, there is the need to investigate and detect misinformation surrounding the fast-moving COVID-19 pandemic, especially through the lens of OSNs.

In response to the current infodemic, this work looks at developing an effective misinformation detection framework with respect to COVID-19, specifically for Twitter. Our major contributions to this area, and the key features of the proposed technique, are summarized as follows:

- We developed a new framework for detecting misinformation on the Twitter platform. In this framework, we integrated six machine-learning algorithms with ensemble learning. Additionally, we enhanced the model by utilizing an ensemble-stacking model with the selected machine-learning models. This resulted in higher accuracy and a more generalized model.
- We collected a large dataset related to the COVID-19 pandemic via Twitter's streaming application program interface (API) for tweets published from 15 January to 15 April, 2020, aiming to explore the features that constitute the detection of misinformation on Twitter.
- The dataset was assessed and labeled by human annotators. After that, we extracted relevant features related to COVID-19 and applied them when building our framework to automatically measure the credibility of the considered tweets.

The rest of this article is organized as follows, Section II presents some works related to misinformation, and its spread, on OSNs. In Section III, we describe our methodology and framework. Section IV presents the results and a performance evaluation. Section V provides more discussions on the results, the learned lessons and the open directions. Finally, we conclude our work in Section VI.

II. RELATED WORK

Misinformation can be defined as inaccurate and false information that is spread knowingly or otherwise [21]. The spread of false information has been occasioned by the improved communication technology that allows many people to share information at an affordable cost [22]. Spreading misinformation can have a negative impact on individual human beings

and the societies in which they live; it can instill fear in people and, by and large, influence their behavioral responses to such things as elections and natural disasters [23]. One such example is the startling statistics that revealed a high prevalence of fake news being spread in the United States, during the months leading up to the 2016 US elections [20], [21]. Another example is the spread of inaccurate information regarding vaccinations, which has been shown to induce phobia in many parents. As a result, some parents with young children have defied medical guidelines and refused to vaccinate or immunize their children [26]. This has resulted in an unprecedented increase in diseases which could otherwise be prevented.

At the societal level, misinformation was ranked by the World Economic Forum in 2013 as a factor that affects financial status and relationships between various countries [27]. The spread of misinformation has risen to greater levels, so that researchers and scholars are taking a keen interest in understanding how misinformation is spread. They have also sought to discover the driving force that motivates people to share false information regarding various events such as the COVID-19 pandemic or political scenarios [28]. For example, Oh *et al.* [29] carried out an analysis of information obtained from Twitter regarding the Haiti earthquake in 2010. They found out that uncertainty and anxiety with regards to information are fundamental factors that determine how false information spreads. They noted that the best way to counter the spread of misinformation is to cite credible information from reliable sources. Domenico *et al.* [30] also sought to establish how scientific misinformation spreads. They concluded that most people deemed information as credible when their friends tweeted about it often.

In [31], the authors investigated the extraction of information from unstructured text in real-time, which is a challenging process. Their work was based on an embedding method for information retrieval from unstructured text corpora. Focusing on the 2014 Ebola and 2016 Zika outbreaks as use cases, the authors collected related data from Twitter and scholarly abstracts from PubMed. Their study validated the robustness of domain-specific word vector models by comparing them against pre-trained generic word vectors. In another work [32], the authors proposed four metrics to obtain a quantitative assessment of the current maturity of social network analysis technologies: pattern & knowledge discovery, information fusion & integration, scalability, and visualization.

More recent research on the spread of fake news was conducted by Vosoughi *et al.* [33], who revealed that misinformation on Twitter was more prevalent than credible news. This phenomenon was associated with the fact that false news contains unfiltered emotions such as fear or even disgust, to which many people can relate, prompting them to share the news with their Twitter friends. Another possible explanation is that people are fond of sharing new information [33]. Zhao *et al.* [34] conducted a study on how misinformation is spread in the Chinese media. They concluded that when a crisis affecting many people occurs, there is a lot of posting and reposting of misinformation on social media. To counter the spread of fake news, the study determined that the acceptable norms of people and their attitude towards sharing fake news were key factors.

In most previous studies that explored the nature and spread of misinformation on social media [35], the focus was on political news and scientific misinformation. However, t here have been very few studies that focused on the spread of misinformation related specifically to COVID-19, which makes it a critical niche to explore.

III. METHODOLOGY

Figure 2 illustrates the system architecture of the proposed framework, starting from collecting the raw data through the use of Twitter's streaming API, towards performance evaluation and comparison. We modelled the COVID-19 misinformation problem as a binary classification problem. Our method is based on a large dataset which was collected and annotated by human experts. The assumption of the methods used in this work is that each tweet used has specific features. In the following subsections, we describe our steps for collecting and labeling the COVID-19-related data from Twitter. Then, we describe the different feature categories used to assess the credibility of each tweet's content.

A. DATA COLLECTION

To conduct out experiment, we created a dataset of tweets that was collected using Twitter's streaming API spanning three months, from 15 January to 15 April, 2020. We used hashtags and keywords related to COVID-19 for our search. Table 1 presents a list of all the relevant keywords used to collect tweets about COVID-19. Identification, annotation, and classification of the tweets were conducted based on the tweet type. In this work, we define a credible COVID-19 tweet as a verified and relevant tweet containing COVID-19 news, vaccination, medication, ways of protection, etc., that had a confirmed, reliable source such as the WHO, IFCN, a fact-checking platform or an official health agency. The non- credible tweets were defined as a collection of non-confirmed or false tweets about COVID-19 that had spread over Twitter. For tweets collected from news segments, detailed research was carried out to determine whether they contained true information. In the case that our annotators determined that a tweet bore false information, it would automatically fall under the ''non-credible'' category.

TABLE 1. COVID-19 Related Keywords.

Keyword	No. of Tweets
Covid-19	411,301
Covid19	545,183
Covid_19	132,396
Coronavirus	32,072
Covid	1,439
Corona	23,411

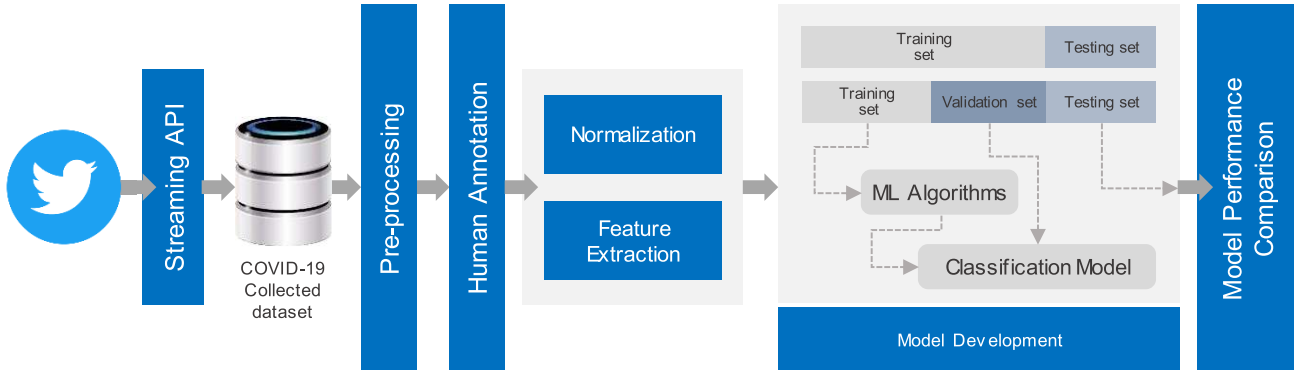

FIGURE 2. System architecture.

Furthermore, to increase the accuracy with which the tweets were categorized, annotators would seek opinions from each other. Ultimately, we compiled a dataset of 980,000 tweets. After deleting duplicated and irrelevant tweets, the final dataset contained more than 400,000 tweets.

Figure 3 illustrates the actual numbers of collected tweets. These tweets were annotated and classified into 121,950 credible tweets and 287,534 non-credible tweets. This dataset forms the primary basis for the subsequent experiments in our work.

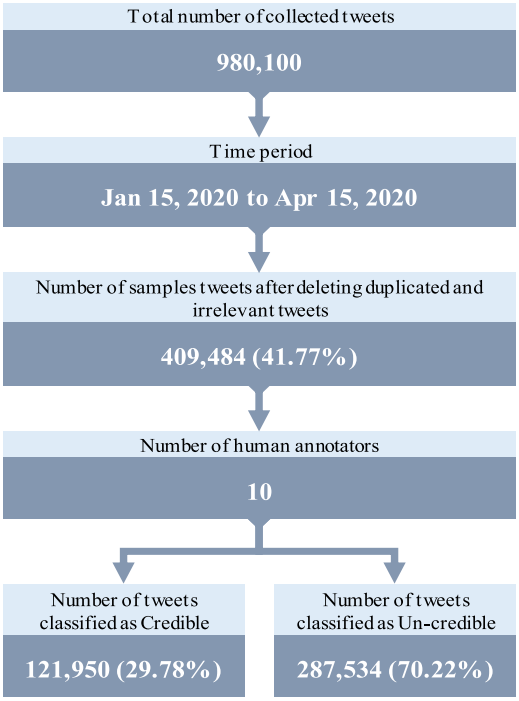

FIGURE 3. COVID-19 collected data.

B. DATA ANNOTATION AND RELIABILITY

The created dataset used in our work was manually annotated by ten human annotators. All the annotators were graduate students with computer and information science majors. Two of the annotators were Ph.D. students, four were master's degree students and the remaining were bachelor's degree students. All the annotators were familiar with the annotation process of OSN content. The annotators agreed to perform with the highest level of accuracy during the annotation process. They worked on a part-time basis, and got paid for three months of work. The average count of annotated tweets per day was about 450 tweets per annotator. Along with the collected tweets regarding COVID-19, we conducted a brief initial meeting with the team to describe the desired dataset, gave annotation guidelines, and provided sources from where the annotators could investigate the credibility of each tweet.

To ensure that the annotators would not be biased towards accounts with high favorite and retweet ratios, we passed the dataset instances to our annotators where the visible features of the users' profiles were omitted. This was done because it is believed that Twitter users with the most followers will likely refrain from spreading misinformation [36].

To evaluate the degree of agreement in the annotation process, and check for inter-agreement between the annotators' classifications, we constructed a small dataset of 20 tweets, and then invited the ten annotators to decide whether the selected tweets were credible or non-credible. We then applied a Bayesian generalized linear mixed model (GLMM) to their results, which is well-suited for measuring agreement between multiple annotators of binary categorical outcomes [37]. The measure of agreement was 0.74, which indicates good reliability. We believe that this strong chance-corrected agreement was due to the clear guidelines of the annotation process and the credible sources provided for checking the credibility of the COVID-19 tweets under study.

C. FEATURE EXTRACTION

The proposed architecture leverages 26 hand-crafted and generic features (described in Table 2) that can be categorized as "tweet-level" and "user-level". Early work on misinformation detection employed supervised-learning techniques, which have been extensively employed to study manually curated features related to content, users, and networks to seek distinguishing features of online credibility [38]. These studies have shown that those features have the potential for distinguishing between credible and non-credible information.

TABLE 2. Description of Extracted Features.

Tweet-level features	
$\text{NoRT}(u_i)$	No. of retweets
$\text{NoUFav}(u_i)$	No. of favorites
$\text{isQM}(t_i)$	Whether the tweet includes a question mark
$\text{isDUP}(t_i)$	Whether the tweet is a duplicate of its source
$\text{isURL}(t_i)$	Whether the tweet contains a URL
$\text{NoURL}(t_i)$	No. of URLs embedded in the tweet
$\text{NoMen}(t_i)$	No. of mentions
$\text{NoHash}(t_i)$	No. of hashtags
$\text{NoWSN}(t_i)$	No. of words in the tweet except the source author's screen name
User-level features	
$\text{NoTwt}(u_i)$	No. of tweets user has posted
$\text{NoPL}(u_i)$	No. of public lists user belongs to
$\text{NoFlw}(u_i)$	No. of followers
$\text{NoFol}(u_i)$	No. of followings
$\text{isBG}(u_i)$	Whether the user profile has a background image
$\text{Rep}(u_i)$	User's reputation (i.e., followers/(followings+1))
$\text{NoTLike}(u_i)$	No. of tweets user has liked so far
$\text{Age}(u_i)$	Account age in days
$\text{IsV}(u_i)$	Whether the profile is verified
$\text{Engage}(u_i)$	User engagement (i.e., # posts / (account age+1))
$\text{FlwR}(u_i)$	Following rate (i.e., followings / (account age+1))
$\text{FavR}(u_i)$	Favorite rate (i.e., user favorites / (account age+1))
$\text{IsGeo}(u_i)$	Whether geolocation is enabled
$\text{IsDesc}(u_i)$	Whether the user has a description
$\text{NoWDesc}(u_i)$	No. of words in the user's description
$\text{NoCharN}(u_i)$	No. of characters in the user's name including white space
$\text{IsSrc}(u_i)$	Whether the user is the source tweet's author

1) USER-LEVEL FEATURES

User-level features represent those features related to the user profile such as number of tweets, number of followers, number of following among others. It is important to mention that some of these features are latent and some are explicitly revealed. For example, the favorite rate is a latent feature, while the number followers is revealed explicitly in the user profile. These features point to the content of the tweet, the user, and the networks. Some works [34]–[36] proved that tweet features curated manually have the capacity to distinguish between credible and non-credible tweets.

2) TWEET-LEVEL FEATURES

In our manual analysis of tweets, we placed a keen focus on content features. We looked at tweets with URLs and embedded multimedia. The purpose of URL usage by Twitter users is to utilize the limited tweeting space [40]. Researchers have shown that many users are likely to share content that has a URL more widely than tweets without [42]. URLs also tend to increase the credibility of the tweet [43]. Notably, it was revealed that tweets that contain URLs go viral more often than ones that do not [44].

D. ENSEMBLE-LEARNING-BASED MODEL

Ensemble learning primarily involves training a group of many single machine-learning models (i.e. weak-learners), and then the results are put together with an ensemble scheme [45]. This enhances the ability of the model to generalize, and also enhances the accuracy levels. Accordingly, we selected a model that had the highest performance among the available models for machine learning.

We keenly analyzed each model based on our dataset and enhanced its parameter tuning. We selected four categories of six models to represent weak-learners in addition to comparing their performances. The four categories include: (1) Naive Bayes (NB) and Bayes net (BN) models, (2) a k-nearest neighbor (kNN) model, (3) a decision Tree (DT) that mainly consisted of C4.5 and random-forest (RF) models, and (4) a support vector machine (SVM). We determined the ability of each of the four categories to detect misinformation, after which we selected a set that included weak-learners in addition to meta-models for ensemble learning.

Ensemble learning exists in multiples strategies. The first is a voting-based model [45]. Here, the model calculates the results of diverse weak-learners, after which it picks the set with the most votes. For a bagging-based model [45], sets of weak-learners get the same weights. Second, for boosting models, the learners have weights that are trained based on data which are misclassified, which is done over a sequence of trained models. As illustrated in Fig. 4, a two-level stacking model was constructed using diverse single machine-learning models.

In this work, we employed commonly used machine-learning techniques to create a stacking-based ensemble-learning model. We carried out various experiments when constructing the ensemble model. For a level-0 weak-learner, we used the SVM+RF models, while for a level-1 weak-learner, we used the C4.5 model as a meta-model. We applied n-fold cross-validation according to the correlation, entropy, and relief when tuning the ultimate model to attain enhanced performance.

IV. EXPERIMENTAL RESULTS

A. EVALUATION OF SINGLE WEAK-LEARNERS

We tested six machine-learning models as candidate weak-learners during the experimental phase: the RF, C4.5, Naive Bayes, Bayes net, SVM and kNN models. Each model was trained based on a 10-fold cross-validation model, which split the overall set into 10 equal subsets. Each category was allocated nine subsets for training and one subset for testing.

A cross-validation model uses the same overall set for training and testing. We took the average accuracy of each model, as well as the standard deviation, which was used as the 1-sigma error, as shown in Fig. 5. It is clear that C4.5 outperformed all the other models. The average accuracy of the cross-validation model for C4.5 was up to 98%, followed by RF, which had an average accuracy of 97.4%. The Naive

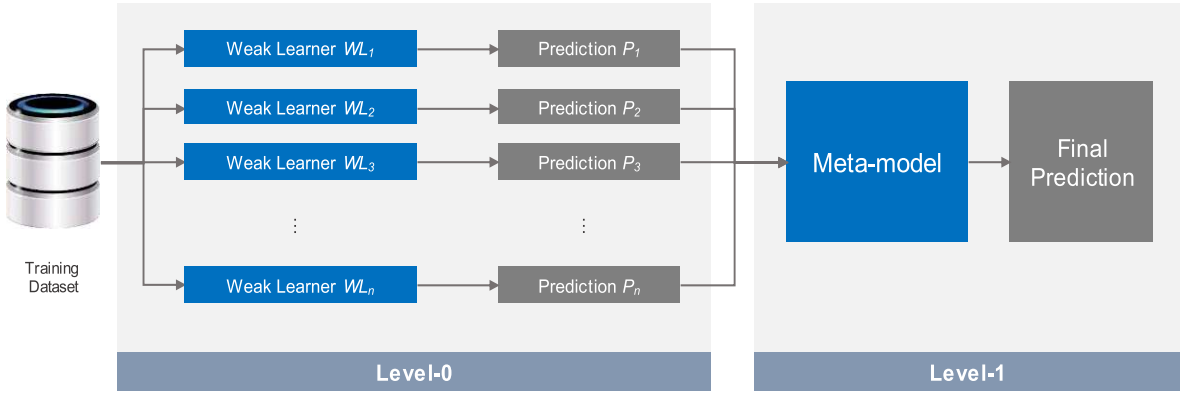

FIGURE 4. Stacking-based ensemble learning with a two-level structure.

Bayes algorithm had the worst performance, with an average accuracy of 85.5%. The results of the average accuracies given for both the cross-validation model and the test set are illustrated in Fig. 5. The standard deviation (error bar) indicates the stability of the selected models.

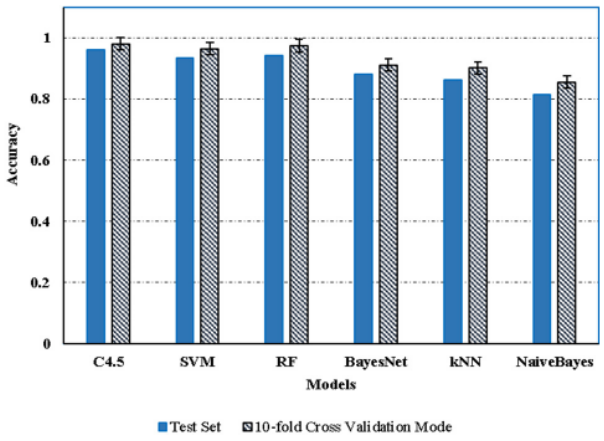

FIGURE 5. Performance comparison of the base models.

The models of the test set showed a slight drop in their accuracies compared with the cross validation model, as seen in Fig. 5. For example, the accuracies of kNN and Naïve Bayes drop significantly (by 4.2 % and 4.05%, respectively), despite the fact they performed the best during the training round. This suggests that these models have an over-fitting issue when training sets are involved. The accuracy of C4.5 only dropped by 2%, and therefore it achieved the highest level of accuracy.

In summary, it was established that a single machine-learning model is likely to be the least efficient when applied to new datasets. This was partly ascertained via the applied learning scheme, and also the quality of the used dataset.

B. STACKING-BASED MODEL

In the preceding section, we analyzed the results from the selected single machine-learning models. It was noted that all models underwent a performance degradation when subjected to the test set. We acknowledge that ensemble learning can possibly transform a single machine-learning model into a robust scheme by integrating several machine-learning models together [40], [41]. However, a fundamental issue of ensemble learning is interpretability losses [46]. Additionally, large models were found to have a vast quantity of parameters in place[47], thus increasing the overfitting problem. A stacking scheme is hence necessary to come up with an acceptable size, which also affects the performance of the given ensemble model and guides the ensemble process. Therefore, we combined models such as level-0 models when performing the stacking. We also replaced the voting module with a meta-model. In this step, the stacking model was used to train the meta-model based on the given available values of the level-0 models.

C. SELECTING THE APPROPRIATE META-MODEL

The stacking model design and the workflow are illustrated in Fig. 4. In this model, a secondary dataset was generated for the meta-model training by use of n-fold cross-validation. Specially, we used 10-fold cross-validation.

The outcome of the meta-data is critical in the final results. We chose to use the selected meta (level-1) models from C4.5, SVM, and kNN. Our choice was made by considering their stable performance when processing new input, as illustrated in Fig. 5, where each of the six given models was applied as a level-0 model. This was done to measure the performance of diverse models combination. Here, the training and testing sets were similar to the ones used in the test experiments presented in Fig. 5. Table 3 presents the results of this particular experiment, which indicate that the performance of three stacked models has a higher average accuracy than the best single-base model when applied to the new test sets.

The kappa statistic [48] was also considered alongside the average accuracy, which is often used to measure the degree of agreement between observed and predicted values for a given dataset. The kappa statistic is defined in Eq. 1, where p is the accuracy of the selected model, and p_r is the accuracy of the random classifier:

$$k = \frac{p - p_r}{1 - p_r} \tag{1}$$

TABLE 3. Meta-models: overall performance.

Measurements	Meta-model		
	C4.5	SVM	kNN
Average Accuracy	95.11%	94.67%	94.39%
Kappa Statistic	0.946	0.939	0.937
RRSE	27.34%	30.72%	31.64%
Weighted Precision	0.953	0.947	0.943
Weighted Recall	0.951	0.946	0.943
Weighted F-score	0.951	0.946	0.943
Weighted AUROC	0.972	0.962	0.963

The kappa statistic has been suggested to be a better measure than the simple accuracy. It has also been established that a higher kappa statistic indicates higher accuracy [49]. The root relative squared error (RRSE) was also considered here, which is presented in Eq. 2, where p_i is the value predicted by the trained model in the ith class (i.e., credible or non-credible), a_i is the actual target rate of the ith class, and $\bar{a}$ is the mean of the training set. In this work, we calculated the total squared error, which we then normalized by dividing it by the overall squared error of the training set.

$$\text{RRSE} = \sqrt{\frac{(p_r - a_1)^2 + \ldots + (p_n - a_n)^2}{(a_1 - \bar{a})^2 + \ldots + (a_n - \bar{a})^2}} \tag{2}$$

The true positive (TP), true negative (TN) and false negative (FN) are used to calculate the model precision, recall and F-score based on the following equations:

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{TN})} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{4}$$

$$\text{F} - \text{score} = \frac{2 \times \text{Percision} \times \text{Recall}}{\text{Percision} + \text{Recall}} \tag{5}$$

The weighted receiver operating characteristic (ROC) area is a probability curve, commonly known as the area under the ROC (AUROC) curve. AUROC uses the weighted TP rate and FP rate values. As shown in Table 4, when applying the SVM+RF combination, the ensemble model achieved better results for all the given measurements.

TABLE 4. Ensemble-models: overall performance.

	Accuracy (%)	Kappa Statistic (%)	RRSE (%)	Weighted AUROC (%)
C4.5+RF	0.973	0.965	0.378	0.994
SVM+RF	**0.978**	**0.975**	**0.257**	**0.997**
C4.5+kNN	0.960	0.948	0.336	0.990
SVM+kNN	0.971	0.962	0.297	0.993
SVM+BayesNet +kNN	0.960	0.948	0.319	0.999
C4.5+Bayes Net+kNN	0.960	0.948	0.328	0.996
All models	0.973	0.965	0.288	1.001

D. SELECTION OF WEAK-LEARNERS

As mentioned earlier, one of our aims was to lower the complexity of the model. As such, we attempted to ascertain the smallest set of base models needed to link all models with high accuracy, while at the same time retaining a simple structure. C4.5 was chosen as the preferred meta-model in the previous subsection, and the following seven combinations of base models were considered: C4.5+RF, SVM+RF, C4.5+kNN, SVM+kNN, SVM+BayesNet+kNN, C4.5+Bayes Net+kNN, and finally all the models together. We used the 10-fold cross-validation model as the stacking model. Our aim was to create training and test sets using the initial training set for a given meta-model, and also to test the complete ensemble model using the test set.

Figure 6 shows the accuracies of the seven different combinations of base models specified above, and Table 4 shows the performance of each model combination. It is clear that the SVM+RF combination has better accuracy (97.8%) than the C4.5 model (96%), although it has lower performance in terms of the remaining three measurements. The SVM+RF combination had greater accuracy compared to the All model. This indicates that combining many models when creating an ensemble model is not necessary as the weak ones may reduce the overall performance. Consequently, a wisely selected meta-model with appropriate combinations of smaller base models may perform better than all the base models put together.

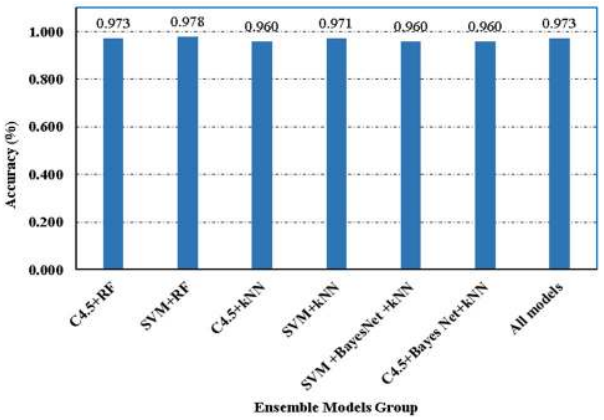

FIGURE 6. Ensemble-models accuracy comparison.

E. SELECTING THE TOP FEATURES

To study the effect of feature importance on model prediction, and given that we have in total 26 features (nine tweet-level features and 17 user-level features), we applied two ranking techniques: entropy-based ranking and correlation-based ranking. Entropy-based ranking was applied in the calculations of the acquired symmetrical data, while correlation-based ranking was applied to sort the features needed depending on their correlation with a given class. Figure 7 represents the calculated values for each of the 26 level features. It can be noted that the top five ranked features were sorted as following: IsV(u_i), NoRT(u_i), NoHash(t_i), NoMen(t_i), and FlwR(u_i). We can conclude that whether an account is verified or not is the most important feature, followed by the number of retweets, number of hashtags, number of mentions and finally, the following rate.

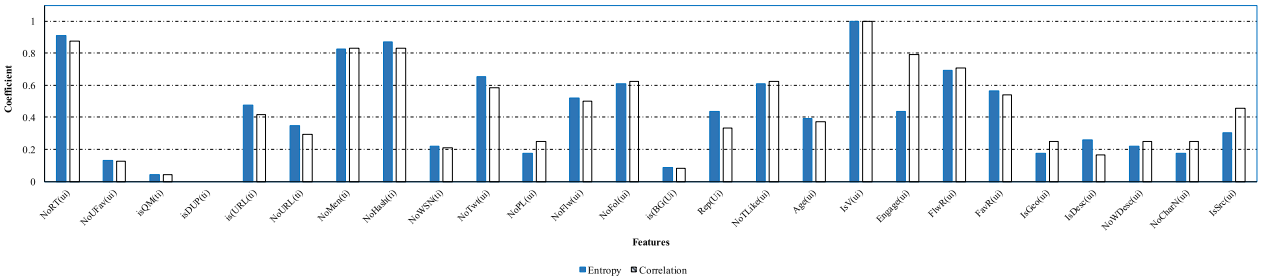

FIGURE 7. Entropy and correlation ranking for the used features.

V. DISCUSSION

In this section, we discuss some lessons which we have learned through conducting this work. Then, we provide a list of open questions. Finally, we discuss some final thoughts that we like to share with the readers during this ongoing crisis.

A. LEARNED LESSONS

Conducting this research revealed several critical issues regarding the use of Twitter data when monitoring the problem of the spread of misinformation during an ongoing crisis. We hope to instigate a discussion of this topic with members of the scientific community. First, we realized that the Twitter API limits the range within which the propagation of misinformation can be researched. Besides, Twitter API does not allow for the quick retrieval of data. Since it takes a while before fact-checking organizations prove or deny a given information, it becomes a challenge to find the public's initial reaction to a given tweet. We also noticed that Twitter opens up COVID-29 data for research purposes [73]. This move, however, does not solve the problem of data retrieval limit.

B. OPEN DIRECTIONS

The spread of misinformation via social media is nothing new, but the current COVID-19 crisis has highlighted the dangers OSNs pose to humanity. This is why there is still a lot of work that needs to be completed regarding the strategies and methods needed to tackle this growing problem. Indeed, the results found here inevitably lead to some key research questions that need to be addressed in the future. We have highlighted four main open research directions in the following.

For starters, there is the need to improve the tools and theories pertaining to social media analytics. This can come in the form of a partially or fully automated way of assessing misinformation. This includes finding the sources of misinformation, the network(s) in which it spreads, how it spreads, how and when it is identified, and the effects such information can have.

Furthermore, since we expect a holistic way of tackling this problem, it is also essential to understand how the results of our work and others in the future can be implemented on other social media platforms, in addition to Twitter. The Twitter network model makes it prone to the spread of misinformation, but other OSNs should also be studied. Indeed, the spread of misinformation regarding COVID-19 is not necessarily restricted to Twitter alone, as it can spread from one platform to another. It is common to see fact-checking platforms failing to refer to any tweets, even in cases where claims are also on Twitter. This is especially the case where the origin of a given claim was on another digital platform, but the sources can be traced back to Twitter, where people might have embedded links from the former OSN. Perhaps a facet that would make an interesting study, in this case, would be closed groups and messages.

The social impact of fake news and misinformation is also an area that requires more research work. Society in general is impacted negatively by the spread of fake news, and studies ought to be conducted to decipher the extent of the impact. We ought to answer questions such as to what extent do the effects occur? Whom do they affect? How can the spread of misinformation be controlled offline? It is also essential to more deeply investigate how the spread of misinformation specifically regarding the COVID-19 pandemic differs from place to place.

Last but not least, there is the need to research this problem in the realms of both social media analytics and crisis management. In this case, social media analytics could borrow from the expertise of managers and researchers in crisis management. This can help them improve their findings and fine-tune their research work. On the flip side, experts in crisis management can use new findings on the propagation of misinformation along with existing models to provide holistic approaches to the way information spreads in a crisis.

C. FINAL THOUGHTS

The spread of misinformation has grown to be a prevalent vice on the Internet over the years. The COVID-19 outbreak has yet again accentuated the risks that come with this growing issue. OSNs have, in particular, been adopted as a way to connect and keep tabs on the information pertaining to the pandemic. However, the spread of misinformation continues to be a major problem as the world continues to grapple with the latest pandemic.

Major international bodies like the WHO and UNESCO are vehemently trying to fight the spread of misinformation related to the COVID-19 outbreak. For example, the WHO has allotted a section of their coronavirus web platform Mythbusters in a bid to tackle fake news. We are, however, in an era of fast-moving information. This begs the question,

how can we fight the spread of misinformation and fake news? Well, finding a solution is undoubtedly easier said than done. We are dealing with an infodemic that has forced top organizations and governments to come up with policies that can curb the spread of fake news and misinformation.

A lot of work towards fighting misinformation has come from academia, but OSNs have also joined in the effort. Top social media platforms have crafted tools, systems, and policies that aim at tackling this issue. However, most have yet to find a tangible solution. Platforms like Twitter and Instagram have had to make changes to handle COVID-19 misinformation among users. Many other OSNs have resorted to giving their own verified updates as a way to deal with this problem. Facebook has also shifted its stance on fake news and misinformation by creating new tools and making policy changes. As per these new policies, the term ''harmful information'' refers to any sort of misinformation that is likely to lead to what is referred to as ''imminent physical harm''. This policy is a response to issues like the spread of messages regarding false cures, and false information regarding things like social distancing, etc.

In such a time when the world is facing a pandemic, it is expected that people could be more responsible with the information they spread. Now is the time where citizens have to adhere to government policies rather than spread unnecessary panic, either directly or even indirectly. Unverified information pertaining to COVID-19 is making the situation worse, and we suggest the following measures as a response to this problem:

1) Users should try and cross-check the information they receive to ascertain credibility. They should check and verify the sources: ideally they should source information from reputable digital platforms like the WHO and other channels.
2) In cases were one may encounter any sort of misinformation, then the best way to control the spread is by avoiding any sort of engagement. In this regard, users should refrain from commenting and sharing such content. This is the only way to stop their growth of popularity.
3) If any form of misinformation is shared on social media, then it is also good to report the content. In this case, if the information was shared privately, you can get in touch with the sender and inform them that content shared is likely misinformation. You can explain the risks associated with such content amidst such a global crisis.
4) One can also contribute to sharing credible content from reputable sources on the web. Use sites and platforms that are known to be credible in relaying information regarding COVID-19.

VI. CONCLUSION

The ongoing COVID-19 pandemic is a threat to human beings. Unlike other global challenges, such as global warming, containing and defeating COVID-19 will depend much on the quality and credibility of information shared amongst people. However, research has shown that misinformation has spread rapidly on OSNs regarding the pandemic.

In this work, we conducted a comprehensive experiment using real data from the Twitter social network. The results revealed that the proposed ensemble-learning model had better performance than single machine-learning-based models. We enhanced the performance of our stacking model by assessing meta-models and weak-learners. We concluded that the final model size can contain fewer features, and it performed slightly better than the original model. As of now, our model has been designed to detect two categories of tweet credibility: credible or non-credible. We acknowledge that there is room for improvement, part of which shall include the following. First is the incorporation of complex tweets containing news and emotional content. Additionally, we shall take into consideration other OSNs that will help us enhance our dataset. Secondly, we shall look out for updated ensemble techniques and machine-learning methods to enhance the current model.

ACKNOWLEDGMENT

The authors thank the Deanship of Scientific Research and RSSU at King Saud University for their technical support.

REFERENCES

[1] C. Sohrabi, Z. Alsafi, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha, ''World health organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19),'' *Int. J. Surg.*, vol. 76, pp. 71–76, Apr. 2020.

[2] M. S. Hossain, G. Muhammad, and N. Guizani, ''Explainable AI and mass surveillance system-based healthcare framework to combat COVID-I9 like pandemics,'' *IEEE Netw.*, vol. 34, no. 4, pp. 126–132, Jul. 2020.

[3] M. A. Rahman, M. S. Hossain, N. A. Alrajeh, and N. Guizani, ''B5G and explainable deep learning assisted healthcare vertical at the edge: COVID-I9 perspective,'' *IEEE Netw.*, vol. 34, no. 4, pp. 98–105, Jul. 2020.

[4] R. Datta, A. K. Yadav, A. Singh, K. Datta, and A. Bansal, ''The infodemics of COVID-19 amongst healthcare professionals in india,'' *Med. J. Armed Forces India*, vol. 76, no. 3, pp. 276–283, Jul. 2020.

[5] A. Mian and S. Khan, ''Coronavirus: The spread of misinformation,'' *BMC Med.*, vol. 18, no. 1, p. 89, Dec. 2020.

[6] N. C. Peeri, N. Shrestha, M. S. Rahman, R. Zaki, Z. Tan, S. Bibi, M. Baghbanzadeh, N. Aghamohammadi, W. Zhang, and U. Haque, ''The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: What lessons have we learned?'' *Int. J. Epidemiol.*, vol. 49, no. 3, pp. 717–726, Jun. 2020.

[7] Poynter Institute. (2020). *The International Fact-Checking Network*. Accessed: Jul. 18, 2020. [Online]. Available: https://www.poynter.org/ifcn/

[8] S. Latif, M. Usman, S. Manzoor, W. Iqbal, J. Qadir, G. Tyson, I. Castro, A. Razi, M. N. K. Boulos, A. Weller, and J. Crowcroft, ''Leveraging data science to combat COVID-19: A comprehensive review,'' *TechRxiv*, 2020, doi: 10.13140/RG.2.2.12685.28644/4.

[9] S. Tasnim, M. M. Hossain, and H. Mazumder, ''Impact of rumors or misinformation on coronavirus disease (COVID-19) in social media,'' *J. Preventive Med. Public Health*, Mar. 2020, doi: 10.31235/osf.io/uf3zn.

[10] M. S. Hossain, G. Muhammad, and A. Alamri, ''Smart healthcare monitoring: A voice pathology detection paradigm for smart cities,'' *Multimedia Syst.*, vol. 25, no. 5, pp. 565–575, Oct. 2019.

[11] R. Kouzy, J. Abi Jaoude, A. Kraitem, M. B. El Alam, B. Karam, E. Adib, J. Zarka, C. Traboulsi, E. Akl, and K. Baddour, ''Coronavirus goes viral: Quantifying the COVID-19 misinformation epidemic on Twitter,'' *Cureus*, vol. 12, no. 3, p. e7255, Mar. 2020.

[12] S. O. Oyeyemi, E. Gabarron, and R. Wynn, ''Ebola, Twitter, and misinformation: A dangerous combination?'' *BMJ*, vol. 349, p. g6178, Oct. 2014.

[13] Y. Ortiz-Martínez and L. F. Jiménez-Arcia, "Yellow fever outbreaks and Twitter: Rumors and misinformation," *Amer. J. Infection Control*, vol. 45, no. 7, pp. 816–817, Jul. 2017.
[14] A. R. Daughton and M. J. Paul, "Identifying protective health behaviors on Twitter: Observational study of travel advisories and Zika virus," *J. Med. Internet Res.*, vol. 21, no. 5, May 2019, Art. no. e13090.
[15] H. Kim and D. Walker, "Leveraging volunteer fact checking to identify misinformation about COVID-19 in social media," *Harvard Kennedy School Misinformation Rev.*, vol. 1, no. 3, pp. 1–10, May 2020, doi: 10.37016/mr-2020-021.
[16] A. Ghenai and Y. Mejova, "Catching zika fever: Application of crowdsourcing and machine learning for tracking health misinformation on Twitter," 2017, *arXiv:1707.03778*. [Online]. Available: http://arxiv.org/abs/1707.03778
[17] H. Rosenberg, S. Syed, and S. Rezaie, "The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic," *CJEM*, vol. 22, no. 4, pp. 418–421, Jul. 2020.
[18] E. Ferrara, "#COVID-19 on Twitter: Bots, conspiracies, and social media activism" 2020, *arXiv:2004.09531*. [Online]. Available: http://arxiv.org/abs/2004.09531
[19] H. W. Park, S. Park, and M. Chong, "Conversations and medical news frames on Twitter: Infodemiological study on COVID-19 in South Korea," *J. Med. Internet Res.*, vol. 22, no. 5, May 2020, Art. no. e18897.
[20] D. Fanelli and F. Piazza, "Analysis and forecast of COVID-19 spreading in China, italy and france," *Chaos, Solitons Fractals*, vol. 134, May 2020, Art. no. 109761.
[21] J. H. Fetzer, "Disinformation: The use of false information," *Minds Mach.*, vol. 14, no. 2, pp. 231–240, May 2004.
[22] X. Chen and S.-C.-J. Sin, "'Misinformation? What of it?' motivations and individual differences in misinformation sharing on social media," *Proc. Amer. Soc. Inf. Sci. Technol.*, vol. 50, no. 1, pp. 1–4, 2013, doi: 10.1002/meet.14505001102.
[23] J. Anderson and L. Rainie, "The future of truth and misinformation online," *Pew Res. Center*, vol. 19, pp. 1–223, Oct. 2017. [Online]. Available: http://www.pewinternet.org/2017/10/19/the-future-of-truth-and-misinformation-online/
[24] M. Carlson, "Fake news as an informational moral panic: The symbolic deviancy of social media during the 2016 US presidential election," *Inf., Commun. Soc.*, vol. 23, no. 3, pp. 374–388, Feb. 2020.
[25] R. R. Mourão and C. T. Robertson, "Fake news as discursive integration: An analysis of sites that publish false, misleading, hyperpartisan and sensational information," *Journalism Stud.*, vol. 20, no. 14, pp. 2077–2095, Oct. 2019.
[26] S. Lewandowsky, U. K. H. Ecker, C. M. Seifert, N. Schwarz, and J. Cook, "Misinformation and its correction: Continued influence and successful debiasing," *Psychol. Sci. Public Interest*, vol. 13, no. 3, pp. 106–131, Dec. 2012.
[27] L. Fernández-Luque and T. Bau, "Health and social media: Perfect storm of information," *Healthcare Informat. Res.*, vol. 21, no. 2, pp. 67–73, 2015.
[28] G. Pennycook, J. McPhetres, Y. Zhang, and D. Rand, "Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy nudge intervention," *Psychol. Sci.*, vol. 31, no. 7, pp. 770–780, 2020.
[29] O. Oh, M. Agrawal, and H. R. Rao, "Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises," *MIS Quart.*, vol. 37, no. 2, pp. 407–426, Feb. 2013.
[30] M. De Domenico, A. Lima, P. Mougel, and M. Musolesi, "The anatomy of a scientific rumor," *Sci. Rep.*, vol. 3, no. 1, p. 2980, Dec. 2013.
[31] A. Khatua, A. Khatua, and E. Cambria, "A tale of two epidemics: Contextual Word2 Vec for classifying Twitter streams during outbreaks," *Inf. Process. Manage.*, vol. 56, no. 1, pp. 247–257, Jan. 2019.
[32] D. Camacho, Á. Panizo-LLedot, G. Bello-Orgaz, A. Gonzalez-Pardo, and E. Cambria, "The four dimensions of social network analysis: An overview of research methods, applications, and software tools," *Inf. Fusion*, vol. 63, pp. 88–120, Nov. 2020.
[33] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018.
[34] L. Zhao, J. Yin, and Y. Song, "An exploration of rumor combating behavior on social media in the context of social crises," *Comput. Hum. Behav.*, vol. 58, pp. 25–36, May 2016.
[35] Q. Fang, J. Sang, C. Xu, and M. S. Hossain, "Relational user attribute inference in social media," *IEEE Trans. Multimedia*, vol. 17, no. 7, pp. 1031–1044, Jul. 2015.
[36] M. Alrubaian, M. Al-Qurishi, M. M. Hassan, and A. Alamri, "A credibility analysis system for assessing information on Twitter," *IEEE Trans. Depend. Sec. Comput.*, vol. 15, no. 4, pp. 661–674, Jul./Aug. 2016.
[37] K. P. Nelson and D. Edwards, "On population-based measures of agreement for binary classifications," *Can. J. Statist.*, vol. 36, no. 3, pp. 411–426, Sep. 2008.
[38] M. Alrubaian, M. Al-Qurishi, A. Alamri, M. Al-Rakhami, M. M. Hassan, and G. Fortino, "Credibility in online social networks: A survey," *IEEE Access*, vol. 7, pp. 2828–2855, 2019.
[39] S. Kwon, M. Cha, and K. Jung, "Rumor detection over varying time windows," *PLoS ONE*, vol. 12, no. 1, Jan. 2017, Art. no. e0168344.
[40] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, "Rumor has it: Identifying misinformation in microblogs," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2011, pp. 1589–1599.
[41] S. Sun, H. Liu, J. He, and X. Du, "Detecting event rumors on Sina Weibo automatically," in *Proc. Asia–Pacific Web Conf.*, 2013, pp. 120–131.
[42] A. Gupta and P. Kumaraguru, "Credibility ranking of tweets during high impact events," in *Proc. 1st Workshop Privacy Secur. Online Social Media (PSOSM)*, 2012, pp. 2–8.
[43] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proc. 20th Int. Conf. World Wide Web (WWW)*, 2011, pp. 675–684.
[44] A. Friggeri, L. Adamic, D. Eckles, and J. Cheng, "Rumor cascades," in *Proc. 8th Int. AAAI Conf. Weblogs Social Media*, 2014, pp. 1–10.
[45] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC Press, 2012.
[46] I. H. Witten and E. Frank, "Data mining: Practical machine learning tools and techniques with Java implementations," *ACM SIGMOD Rec.*, vol. 31, no. 1, pp. 76–77, Mar. 2002.
[47] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
[48] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: The kappa statistic," *Family Med.*, vol. 37, no. 5, pp. 360–363, 2005.
[49] A. Bendavid, "Comparison of classification accuracy using Cohen's weighted kappa," *Expert Syst. Appl.*, vol. 34, no. 2, pp. 825–832, Feb. 2008.

MABROOK S. AL-RAKHAMI (Member, IEEE) received the master's degree in information systems from King Saud University, Riyadh, Saudi Arabia, where he is currently pursuing the Ph.D. degree with the Information Systems Department, College of Computer and Information Sciences. He has worked as a Lecturer and taught many courses, such as programming languages in computer and information science, with King Saud University, Muzahimiyah Branch. He has authored several papers in peer-reviewed IEEE/ACM/Springer/Wiley journals and conferences. His research interests include edge intelligence, social networks, cloud computing, the Internet of Things, big data, and health informatics.

ATIF M. AL-AMRI (Member, IEEE) received the Ph.D. degree in computer science from the School of Information Technology and Engineering, University of Ottawa, Canada, in 2010. He is currently a Full Professor with the Software Engineering Department, College of Computer and Information Sciences (CCIS), King Saud University (KSU), Riyadh, Saudi Arabia. He is one of the founding members of the Chair of Pervasive and Mobile Computing (CPMC), CCIS, KSU, successfully managing its research program, which transformed the chair as one of the best chairs of research excellence in the college. His research interests include multimedia-assisted health systems, ambient intelligence, service-oriented architecture, multimedia cloud, sensor cloud, the Internet of Things, big data, mobile cloud, social networks, and recommender systems.