Detecting Health-Related Rumors on Twitter using Machine Learning Methods

Faisal Saeed[1], Mohammed Al-Sarem*[3]
Essa Abdullah Hezzam[4]
Department of Information System
College of Computer Science and Engineering,
Taibah University, Medina, Saudi Arabia

Wael M.S. Yafooz[2]
Computer Science Department
College of Computer Science and Engineering,
Taibah University, Medina, Saudi Arabia

***Abstract*—Nowadays, the huge usage of internet leads to tremendous information growth as a result of our daily activities that deal with different sources such as news articles, forums, websites, emails and social media. Social media is a rich source of information that deeply affect users by its useful content. However, there are a lot of rumors in these social media platforms which can cause critical consequences to the people's lives, especially if it is related to the health-related information. Several studies focused on automatically detecting rumors from social media by applying machine learning and intelligent methods. However, few studies concerned about health-related rumors in Arabic language. Therefore, this paper is dealing with detecting health-related rumors focusing on cancer treatment information that are spread over social media using Arabic language. In addition, it presents the process of creating a dataset that is called Health-Related Rumors Dataset (HRRD) which will be available and beneficial for further studies in health-related research. Furthermore, an experiment has been conducted to investigate the performance of several machine learning methods to detect the health-related rumors on social media for Arabic language. The experimental results showed the rumors can be detected with an accuracy of 83.50%.**

Keywords*—Health-related misinformation; cancer disease; fake information; Twitter; classification formatting

I. Introduction

Tremendous amount of information are generated as a result of our daily activities and from different sources such as news articles, forum, websites, emails and social media. Therefore, information spread quickly, especially through social media such as Facebook, Twitter, Instagram and others. Hence, social media is a rich source of information that deeply affect users by its contents. This content can be useful for the needs of many users in different areas such education, politics, economics, advertisement, health care, shopping and others. At the same time, there are a lot of information which can be false (rumor) [1][2][3].

Social networks are established in order to connecting people, enhancing relationship and sharing useful information [4]. Recently, it becomes the communication channel for education, advertisement and many other activities. There are a lot of benefits of it in marketing [5] [6], and other professional purposes [7] [8]. In despite of advantageous provided from social networks, the quality of information is low, especially on news and health care information [1] and [2]. Thus, anyone using a social media is able to write self-content as advice or recommendation even without a-prior knowledge and spread such information to many people in minutes [9]. This information could be related to medical treatment and health-related issues [2] [10] [11] [12]. In addition, social media users widely rely on themselves to obtain medical advices from social media. Therefore, the creditability of such information is very important.

Information on social media lacks to quality, credibility and trust-ability as emphasized in health-related misinformation [13] [14] [15]. This misinformation/rumors could have critical consequences to the people's life, especially if it concerns on health information that can lead to health risks [16]. In the existing studies of detecting rumors in health-related information, a little attention has been given to cancer-related information using Arabic language. Therefore, the purpose of this paper is to apply several machine learning methods for detecting health-related rumors aiming cancer treatment over social media using Arabic language. In addition, a dataset for cancer information treatment called Health-Related Rumors Dataset (HRRD) has been created. HRRD has been collected from Twitter, classify by domain experts into true and false information. Then, different preprocessing methods were applied on the dataset such as stemming, tokenization, feature extraction and oversampling. Furthermore, several machine learning methods have been applied and evaluated using different metrics.

This paper organizes as follows: Section II demonstrates related studies. The methods and materials are presented in Section III. While, Section IV explains the results and discussion. Finally, conclusion of this paper is highlighted in Section V.

II. Related Works

There are few studies focusing on rumors on social media such as [17] [18] [19] [20] [21] [22], while limited studies were conducted to detect rumors about health-related information. Such studies can be classified into focusing on correctness and reliability of medical precipitations [23] [24] [25] effects global health and health literacy [13] [14] [15] [26] [27], and detecting health-related rumors [28] [29] [30].

Soon et al. [23] and Zhang et al. [28] highlighted the issue of health-related rumors by identifying the consequences and benefits of the personal's perceptions through online platform. While, the studies in [13] [14] [15] [26] [27] [31] emphasized

*Corresponding Author

on the importance of checking and verifying online perceptions and credibility of information by health professionals and physicians in domain, otherwise it will be harmful to user's health. In the other way, the authors in [29] [30] [32] presented rumor detection methods by detecting the health-related misinformation using extracting and identifying the fake features. In [30] and [32], Health-related Misinformation Detection framework was developed in order to detect unreliable and reliable health-related information.

Zhang et al. [33] applied logistic regression model for distinguishing between true and false health-related rumors. For this purpose, 453 health rumors from Chinese website were collected and analyzed. The results showed that lengths of rumors headlines, statements and presence of pictures within the context are the most distinctive indicators of false rumors, whereas rumors that contain numbers, hyperlinks and source cues are more likely to be true.

On the other hand, the authors in [34] studied human behavior regarding travel to these areas affected by Zika virus. They have combined content analysis with several machine learning techniques in order to identify first-person reactions and change of travel-related decisions during the Zika outbreak. For this purpose, 29,386 English-language tweets were collected. Only 2000 English-language tweets were annotated and labeled by two annotators and then out of them, 400 tweets were used for training binary logistic regression classifier. The classifier's performance was evaluated using Precision, Recall and F1-score. The best F1-scores were 0.65 for travel change decision, 0.63 for travel consideration and 0.92 for identifying the first-person reactions. In [35], it was dealt with the Zika virus outbreak and gathered around 30 million tweets posted around the world. They incorporated health professionals and crowdsourcing methods to capture and annotated health-related rumors, and used several machine learning techniques including naïve Bayes, random forest and random decision tree to classify the tweets. The data set consists of 3,343 labeled tweets, in which 1786 were rumors. Regarding to the performance of the classifiers, the best achieved results were yielded when random tree was employed with precision of 0.946 and recall 0.944. In [36], they examined questionable health-related information that are posting on Twitter, in particular these tweets related to cancer treatments. For this purpose, they studied 3,212 Twitter users who posted unverified information about cancer treatment. A total of 215,109 tweets about rumor topics were harvested. Then, rigorous filter criteria were applied to exclude irrelevant tweets and users accounts from the data set. At the end, only 4,000 tweets remained, total of which 2,890 were labeled as information about cancer and 1,110 tweets were labeled as non-related to the cancer topic. The logistic regression using n-grams features was employed on this dataset and showed good results.

In addition, the authors in [37] examined 1.5 million tweets mentioning obesity and diabetes epidemics. The main purpose of this study was to assess the quality of information circulating in these conversations, as well as the behavior and information needs of the users engaged in it. The results showed that 41% of the circulated obesity-related tweets and 50% diabetes were posted by non-governmental or academic institution. Furthermore, other studies focused on creating automatically a health misinformation dataset harvested from an online health discussion forum such as [38]. Also, [39] analyzed vaccine rumors in news and social media by developing a dashboard platform that has two networks visualization: the user-as-nodes and tweets-as-nodes. To demonstrate the robustness of the system, a total of 875,088 tweets and 4,020 news articles about vaccine-related topics were collected. It was found that this tool is useful only for tracking the most influenced accounts who post frequently such news or tweets. Similarly, [40] modeled the trustworthiness and reliability of online information using deep learning technique, in particular, convolutional neural network (CNN). The applied model was used to generate a recommendation of trusted medical articles with average veracity score of 78.32%.

III. Materials and Methods

The main methodology of this study is illustrated in Fig. 1. It briefly shows the main four phases of conducting this study, which include dataset generation, data preprocessing, applying machine learning methods and evaluation the model.

A. Phase 1: Dataset Generation

The dataset generated for this study is called Health-Related Rumors Dataset (HRRD), which includes a collection of tweets that are related to rumors on cancer disease/treatment. To the best of our knowledge, there is no available dataset for health-related rumors on social media for Arabic language. The phases of generating this dataset includes five steps, which are identifying the keywords, extracting the tweets using Twitter search APIs, extracting the tweets manually, screening the tweets, and labeling the tweets. In this study, the collected tweets are related to cancer symptoms, causes, prevention, treatment and awareness on tweeter that were written in Arabic language. These five steps are described in detail as follows:

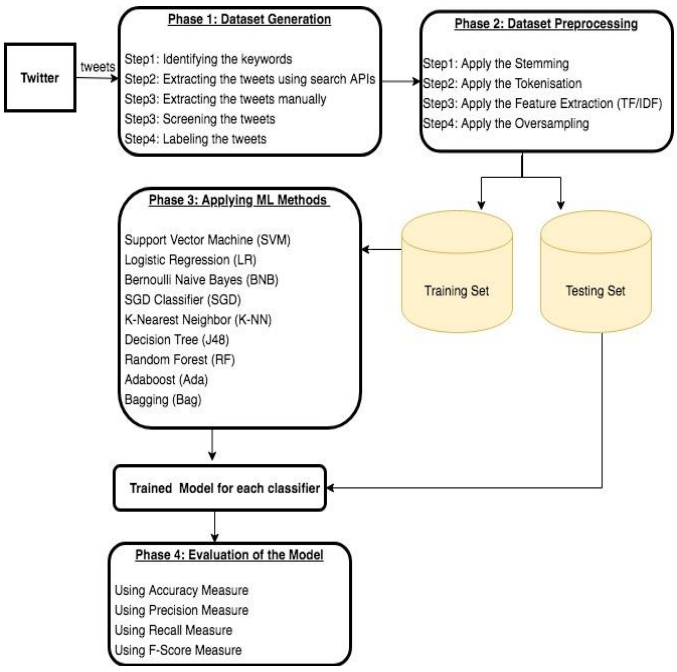

Fig. 1. Methodology of the Study.

Step 1: Identifying the keywords

Several keywords were used in order to automatically or manually extract tweets regarding to cancer disease in Arabic language. These include: cancer disease, cancer causes, cancer treatment, fighting cancer, awareness about cancer, campaign about cancer, warning about cancer, health and cancer, avoid cancer and information about cancer.

Step 2: Extracting the tweets using Twitter search APIs

Using the keywords mentioned in the previous step, 18,684 tweets were automatically extracted from Twitter using Twitter search APIs. These APIs are based on the REST architecture which allow to access Twitter data such as tweets and the user profile information. However, due to the limitation on tweets extraction, the user can only perform a limited number of requests daily. Therefore, additional tweets were extracted manually, as described in the next step.

Step 3: Extracting the tweets manually

In this step, tweets extracted manually due to limitations of extracting tweets using Twitter search APIs which also retrieve huge number of irrelevant tweets, additional tweets were extracted manually using the above keywords to provide more relevant tweets to the dataset. A total of 180 tweets were extracted manually.

Step 4: Screening the tweets

The extracted tweets in step 1 and 2 were screened manually to exclude any irrelevant tweets, which were posted by product sellers, companies, fake/untrusted accounts and others. The total number of tweets was reduced tweets to 175 tweets.

Step 5: Labeling the tweets

In the process of labeling, the extracted tweets were divided into four groups and sent to domain experts (medical doctors) to label the tweets into three options: rumors= yes, no and not sure. The first group were answered by nine experts, while the rest were answered by seven experts only. The majority voting was used to find the final label for each tweet. The results of labeling were (yes: 31, no: 41, 87: not sure, 16: the decision cannot be made because of equal voting). Then, the tweets with labels: "not sure" and "no label" were combined (103) for re-labeling. Also, additional tweets were extracted manually and included to this group (33 tweets). These tweets were divided into three groups and sent to domain experts including oncologists. The first group was answered by seven experts, the second group was answered by four experts and the last group was answered by two experts. The majority voting also was applied here to combine the votes and label the data. In case of any not-sure answers or equal number of votes, more weights were given to the oncologist's answers. The total number of labeled tweets were 208, which include: yes: 128, no: 80. The distribution of the classes for the final dataset is shown in Fig. 2.

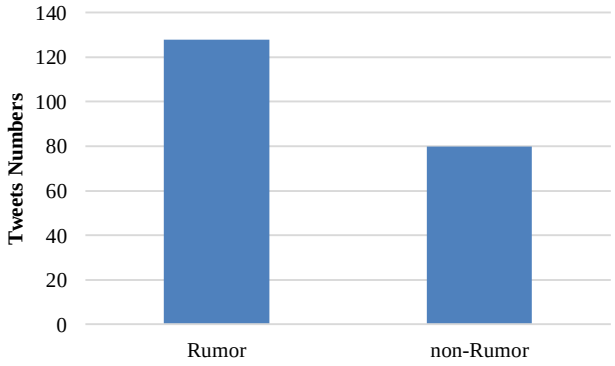

Fig. 2. Distribution of Classes in HRRD Dataset.

B. Phase 2: Data Preprocessing

Python 3.6 with Windows 10 operating system were used for preprocessing the dataset and conduct the experiments. Several libraries were installed including: NLKT for stemming of Arabic texts. In addition, the raw data were tokenized and represented using unigram, bi-gram, trigram, 4-gram and 5-gram. The feature extraction was performed using TF-IDF. The impact of these different preprocessing methods was investigated for detection of health-related rumors in Arabic language. In addition, as shown in Fig. 2, the dataset is unbalanced. Therefore, oversampling method was applied on the minority class in order to provide balanced dataset. The impact of oversampling method also was investigated.

C. Phase 3: Machine Learning Methods

To detect the health-related rumors for Arabic language in social media (Twitter), several machine learning methods were used which include Support Vector Machine (SVM), Logistic Regression (LR), Bernoulli Naive Bayes (BNB), SGD Classifier, K- Nearest Neighbor (K-NN) and Decision Tree (J48). In addition, three ensemble machine methods were used which are Random Forest (RF), AdaBoost (Ada), and Bagging (Bag).

To apply the machine learning methods, the dataset was split into training set (70%) and testing set (30%). Then, several evaluation metrics were applied to measure the performance of detecting the health-related rumors for Arabic language in social media. The details of these measures are described in the next subsection.

D. Evacuation Metric

The evaluation metrics were used for evaluating classification methods that were combined with different preprocessing methods. These includes: Precision, Recall, F-1 score and Accuracy. The definition of these measurements is illustrated as follows:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1\ score = \frac{2*(Precision*Recall)}{Precision+Recall} \quad (3)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

where TP is true positive; TN is true negative; FP is false positive, and FN is false negative.

IV. Experimental Results and Discussion

The experiments have been conducted on two stages *without and with applying oversampling* for the dataset. In each stage, five steps were done by applying different tokenization methods (1-5 gram). The accuracy of each classifier was reported, and the precision, recall and F-score values for rumor class and for the two classes (weighted value) were presented. The best value(s) of each evaluation criterion was highlighted (bold).

Tables I to V shows the performance of the nine machine learning methods for different tokenization methods (1-5 gram) before applying the oversampling. The results showed that the best accuracy was obtained using SGD classifier (76.19%) for bigram method. For detecting the rumor class, the best precision (0.80) was obtained by Adaboost classifier with 5-gram method, while both BNB and RF obtained the best recall values (1.0) using all tokenization methods. Furthermore, the best F-score obtained for this class was (0.82) by SGD classifier with bigram method. On the other hand, the best weighted precision (for the two classes) was obtained by KNN classifier (0.80) with trigram method, while the best recall and F-score values were obtained using SGD classifier (0.76, 0.75 respectively) with bigram. The experiments reported the good performance of SGD classifier to detect health-related rumors using unbalanced dataset (before oversampling).

TABLE I. The Performance of Machine Learning Methods for Detecting Health-Related Rumors (without Oversampling and using Unigram)

Classifier	Acc.	For Rumor Class			For Two classes (Weighted Avg.)		
		Precision	*Recall*	*F-score*	*Precision*	*Recall*	*F- Score*
SVM	**74.60%**	**0.76**	0.81	0.78	0.74	**0.75**	**0.74**
LR	**74.60%**	0.75	0.83	**0.79**	0.75	**0.75**	**0.74**
BNB	66.67%	0.63	**1.00**	0.77	**0.79**	0.67	0.60
RF	57.14%	0.57	**1.00**	0.73	0.33	0.57	0.42
SGD	**74.60%**	0.75	0.83	**0.79**	0.75	**0.75**	**0.74**
KNN	61.90%	0.62	0.86	0.72	0.62	0.62	0.58
J48	63.49%	0.57	0.59	0.58	0.63	0.63	0.63
Ada	68.25%	0.70	0.78	0.74	0.68	0.68	0.68
Bag	66.66%	0.67	0.81	0.73	0.66	0.67	0.66

TABLE II. The performance of machine learning methods for detecting health-related rumors (without oversampling and using bigram)

Classifier	Acc.	For Rumor Class			For Two classes (Weighted Avg.)		
		Precision	*Recall*	*F-score*	*Precision*	*Recall*	*F- Score*
SVM	71.43%	0.72	0.81	0.76	0.71	**0.71**	0.71
LR	71.43%	0.71	0.83	0.77	0.71	**0.71**	0.71
BNB	65.08%	0.62	**1.00**	0.77	0.78	0.65	0.57
RF	57.14%	0.57	**1.00**	0.73	0.33	0.57	0.42
SGD	**76.19%**	0.72	0.94	**0.82**	**0.79**	**0.76**	**0.75**
KNN	71.43%	0.67	0.97	0.80	0.77	0.71	0.68
J48	65.08%	0.65	0.83	0.73	0.65	0.65	0.63
Ada	63.49%	0.71	0.61	0.66	0.65	0.63	0.64
Bag	69.84%	**0.73**	0.75	0.74	0.70	0.70	0.70

TABLE III. The performance of machine learning methods for detecting health-related rumors (without oversampling and using trigram)

Classifier	Acc.	For Rumor Class			For Two classes (Weighted Avg.)		
		Precision	*Recall*	*F-score*	*Precision*	*Recall*	*F- Score*
SVM	**73.02%**	**0.72**	0.86	0.78	0.73	0.73	**0.72**
LR	71.43%	0.70	0.89	0.78	0.73	0.71	0.70
BNB	65.08%	0.62	**1.00**	0.77	0.78	0.65	0.57
RF	57.14%	0.57	**1.00**	0.73	0.33	0.57	0.42
SGD	68.25%	0.67	0.89	0.76	0.70	0.68	0.66
KNN	74.60%	0.70	0.97	**0.81**	**0.80**	**0.75**	**0.72**
J48	68.25%	**0.72**	0.72	0.72	0.68	0.68	0.68
Ada	57.14%	0.65	0.56	0.60	0.58	0.57	0.57
Bag	68.25%	0.71	0.75	0.73	0.68	0.68	0.68

TABLE IV. THE PERFORMANCE OF MACHINE LEARNING METHODS FOR DETECTING HEALTH-RELATED RUMORS (WITHOUT OVERSAMPLING AND USING 4-GRAM)

Classifier	Acc.	For Rumor Class			For Two classes (Weighted Avg.)		
		Precision	*Recall*	*F-score*	*Precision*	*Recall*	*F- Score*
SVM	**73.02%**	0.71	0.89	0.79	0.74	**0.73**	**0.72**
LR	71.43%	0.70	0.89	0.78	0.73	0.71	0.70
BNB	65.08%	0.62	**1.00**	0.77	**0.78**	0.65	0.57
RF	57.14%	0.57	**1.00**	0.73	0.33	0.57	0.42
SGD	69.84%	0.67	0.94	0.78	0.74	0.70	0.67
KNN	71.43%	0.67	0.97	**0.80**	0.77	0.71	0.68
J48	58.73%	0.68	0.53	0.59	0.61	0.59	0.59
Ada	71.43%	**0.78**	0.69	0.74	0.72	0.71	**0.72**
Bag	71.43%	0.74	0.78	0.76	0.71	0.71	0.71

TABLE V. THE PERFORMANCE OF MACHINE LEARNING METHODS FOR DETECTING HEALTH-RELATED RUMORS (WITHOUT OVERSAMPLING AND USING 5-GRAM)

Classifier	Acc.	For Rumor Class			For Two classes (Weighted Avg.)		
		Precision	*Recall*	*F-score*	*Precision*	*Recall*	*F- Score*
SVM	**71.43%**	0.70	0.89	0.78	0.73	**0.71**	**0.70**
LR	69.84%	0.68	0.89	0.77	0.71	0.70	0.68
BNB	63.49%	0.61	**1.00**	0.76	**0.78**	0.63	0.54
RF	57.14%	0.57	**1.00**	0.73	0.33	0.57	0.42
SGD	69.84%	0.67	0.94	0.78	0.74	0.70	0.67
KNN	69.84%	0.66	0.97	**0.79**	0.76	0.70	0.66
J48	52.38%	0.62	0.42	0.50	0.55	0.52	0.52
Ada	61.90%	**0.80**	0.44	0.57	0.69	0.62	0.61
Bag	61.90%	0.67	0.67	0.67	0.62	0.62	0.62

In the second stage, oversampling method was applied and the five tokenization methods (1-5 gram) were used. The performance of detecting the health-related rumors using machine learning was consistently improved. Tables VI to X show the performance of the nine machine learning methods with oversampling method. The results showed that the best accuracy was obtained by RF classifier (83.50%) with 4 and 5-gram, and by using SGD classifier (83.50%) using 4-gram method.

For detecting the rumor class, the best precision value (0.83) was obtained by Bag (with unigram) and LR (with 4 and 5-gram), while the best recall and F-score values (1.0 and 0.86, respectively) was obtained by BNB classifier using 3, 4 and 5 -gram).

For the weighted average values, the best precision and recall value was obtained by BNB (0.87, 0.83 respectively) with trigram method. In addition, other classifiers obtained superior performance for recall such as SGD and RF classifiers. The best F-score value was obtained by RF (0.83) using 4 and 5–gram methods.

To compare the accuracy of all machine learning methods using all tokenization methods with and without oversampling, the results were summarized in Fig. 3 to 7. The results showed the consistent enhancements obtained when oversampling was used for all machine learning methods. The best accuracy was obtained by RF using 4 and 5-grams.

TABLE VI. THE PERFORMANCE OF MACHINE LEARNING METHODS FOR DETECTING HEALTH-RELATED RUMORS (WITH OVERSAMPLING AND USING UNIGRAM)

Classifier	Acc.	For Rumor Class			For Two classes (Weighted Avg.)		
		Precision	*Recall*	*F-score*	*Precision*	*Recall*	*F- Score*
SVM	71.84%	0.79	0.63	0.70	0.73	0.72	0.72
LR	73.79%	0.81	0.65	0.72	0.75	0.74	0.74
BNB	79.61%	0.74	**0.94**	**0.83**	**0.82**	0.80	0.79
RF	**80.58%**	0.79	0.85	0.82	0.81	**0.81**	**0.81**
SGD	71.84%	0.80	0.61	0.69	0.74	0.72	0.72
KNN	57.28%	0.63	0.44	0.52	0.59	0.57	0.57
J48	79.61%	**0.84**	0.76	0.80	0.80	0.80	0.80
Ada	76.70%	0.76	0.81	0.79	0.77	0.77	0.77
Bag	**80.58%**	0.83	0.80	0.81	0.81	**0.81**	**0.81**

TABLE VII. THE PERFORMANCE OF MACHINE LEARNING METHODS FOR DETECTING HEALTH-RELATED RUMORS (WITH OVERSAMPLING AND USING BIGRAM)

Classifier	Acc.	For Rumor Class			For Two classes (Weighted Avg.)		
		Precision	*Recall*	*F-score*	*Precision*	*Recall*	*F- Score*
SVM	74.76%	**0.82**	0.67	0.73	0.76	0.75	0.75
LR	75.73%	**0.82**	0.69	0.75	0.77	0.76	0.76
BNB	**81.55%**	0.75	**0.98**	**0.85**	**0.85**	**0.82**	**0.81**
RF	**81.55%**	0.77	0.93	0.84	0.83	**0.82**	**0.81**
SGD	78.64%	0.79	0.81	0.80	0.79	0.79	0.79
KNN	64.08%	0.76	0.46	0.57	0.68	0.64	0.63
J48	76.70%	0.77	0.80	0.78	0.77	0.77	0.77
Ada	78.64%	0.78	0.83	0.80	0.79	0.79	0.79
Bag	78.64%	0.78	0.83	0.80	0.79	0.79	0.79

TABLE VIII. THE PERFORMANCE OF MACHINE LEARNING METHODS FOR DETECTING HEALTH-RELATED RUMORS (WITH OVERSAMPLING AND USING TRIGRAM)

Classifier	Acc.	For Rumor Class			For Two classes (Weighted Avg.)		
		Precision	*Recall*	*F-score*	*Precision*	*Recall*	*F- Score*
SVM	74.76%	**0.82**	0.67	0.73	0.76	0.75	0.75
LR	75.73%	**0.82**	0.69	0.75	0.77	0.76	0.76
BNB	**82.52%**	0.75	**1.00**	**0.86**	**0.87**	**0.83**	**0.82**
RF	80.58%	0.75	0.94	0.84	0.83	0.81	0.80
SGD	**82.52%**	0.80	0.89	0.84	0.83	**0.83**	**0.82**
KNN	62.14%	0.73	0.44	0.55	0.65	0.62	0.61
J48	80.58%	0.76	0.93	0.83	0.82	0.81	0.80
Ada	74.76%	0.79	0.70	0.75	0.75	0.75	0.75
Bag	76.70%	0.78	0.78	0.78	0.77	0.77	0.77

TABLE IX. THE PERFORMANCE OF MACHINE LEARNING METHODS FOR DETECTING HEALTH-RELATED RUMORS (WITH OVERSAMPLING AND USING 4-GRAM)

Classifier	Acc.	For Rumor Class			For Two classes (Weighted Avg.)		
		Precision	*Recall*	*F-score*	*Precision*	*Recall*	*F- Score*
SVM	75.73%	0.82	0.69	0.75	0.77	0.76	0.76
LR	76.70%	**0.83**	0.70	0.76	0.78	0.77	0.77
BNB	81.55%	0.74	**1.00**	0.85	**0.86**	0.82	0.81
RF	**83.50%**	0.77	0.98	**0.86**	**0.86**	**0.83**	**0.83**
SGD	**83.50%**	0.79	0.93	0.85	0.84	**0.83**	**0.83**
KNN	63.11%	0.74	0.46	0.57	0.66	0.63	0.62
J48	67.96%	0.72	0.63	0.67	0.69	0.68	0.68
Ada	79.61%	0.78	0.85	0.81	0.80	0.80	0.80
Bag	73.79%	0.76	0.72	0.74	0.74	0.74	0.74

TABLE X. THE PERFORMANCE OF MACHINE LEARNING METHODS FOR DETECTING HEALTH-RELATED RUMORS (WITH OVERSAMPLING AND USING 5-GRAM)

Classifier	Acc.	For Rumor Class			For Two classes (Weighted Avg.)		
		Precision	*Recall*	*F-score*	*Precision*	*Recall*	*F- Score*
SVM	75.73%	0.82	0.69	0.75	0.77	0.76	0.76
LR	77.67%	**0.83**	0.72	0.77	0.78	0.78	0.78
BNB	81.55%	0.74	**1.00**	0.85	**0.86**	0.82	0.81
RF	**83.50%**	0.77	0.98	**0.86**	**0.86**	**0.83**	**0.83**
SGD	79.61%	0.79	0.83	0.81	0.80	0.80	0.80
KNN	63.11%	0.74	0.46	0.57	0.66	0.63	0.62
J48	78.64%	0.76	0.87	0.81	0.79	0.79	0.78
Ada	77.67%	0.76	0.83	0.80	0.78	0.78	0.78
Bag	74.76%	0.73	0.81	0.77	0.75	0.75	0.75

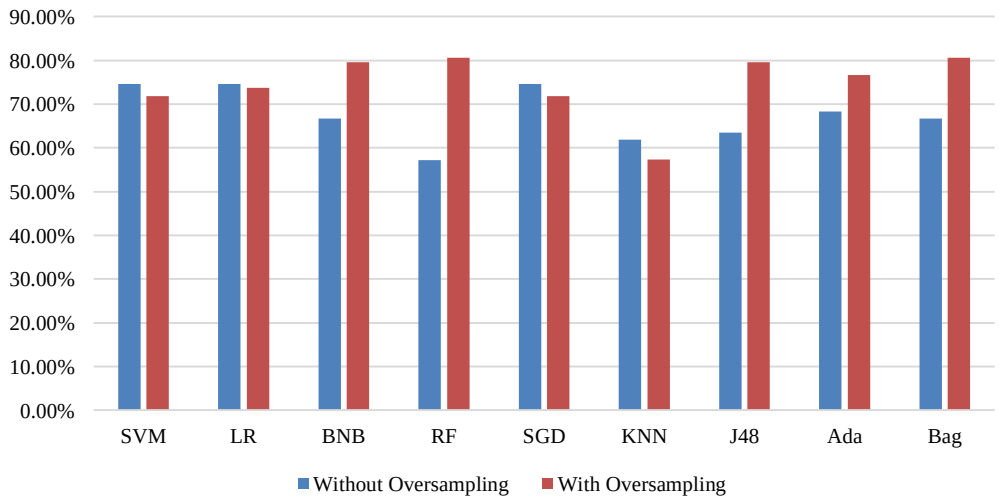

Fig. 3. Accuracy of Classifiers with Unigram.

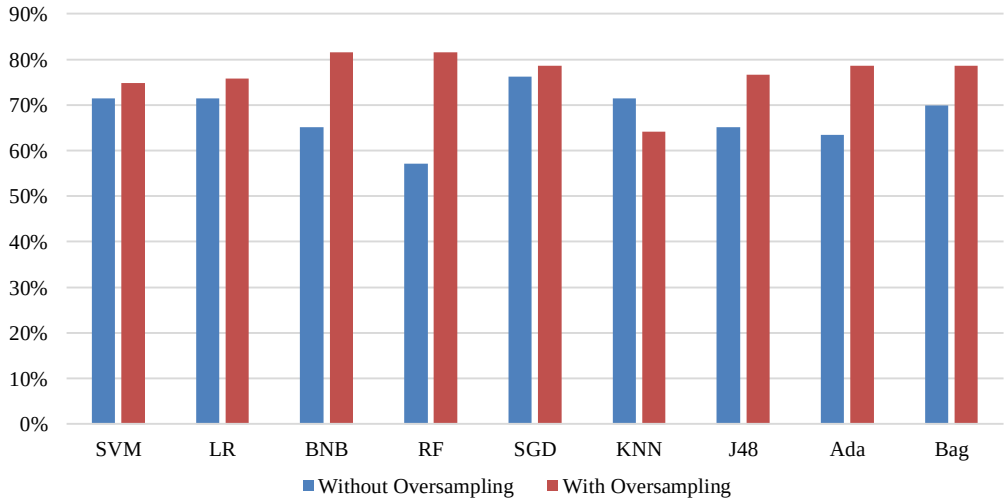

Fig. 4. Accuracy of Classifiers with Bigram.

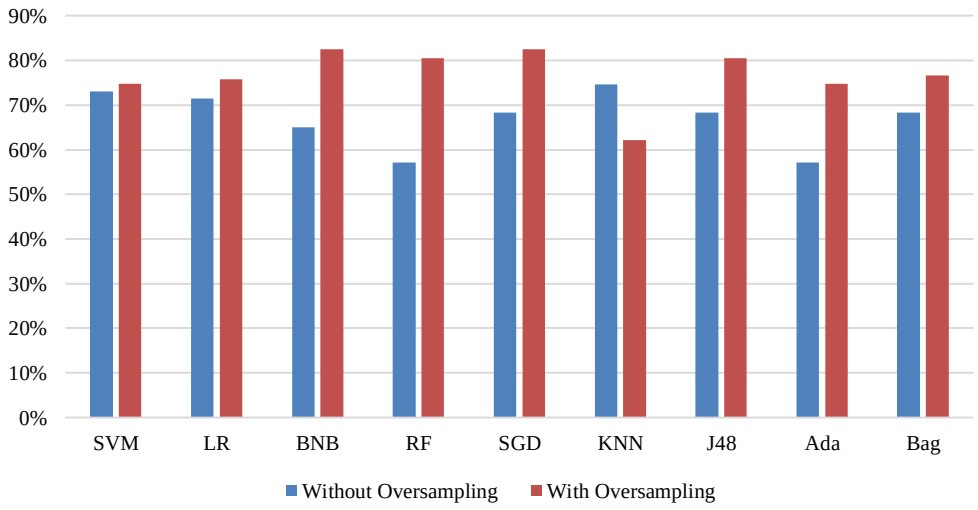

Fig. 5. Accuracy of Classifiers with Trigram.

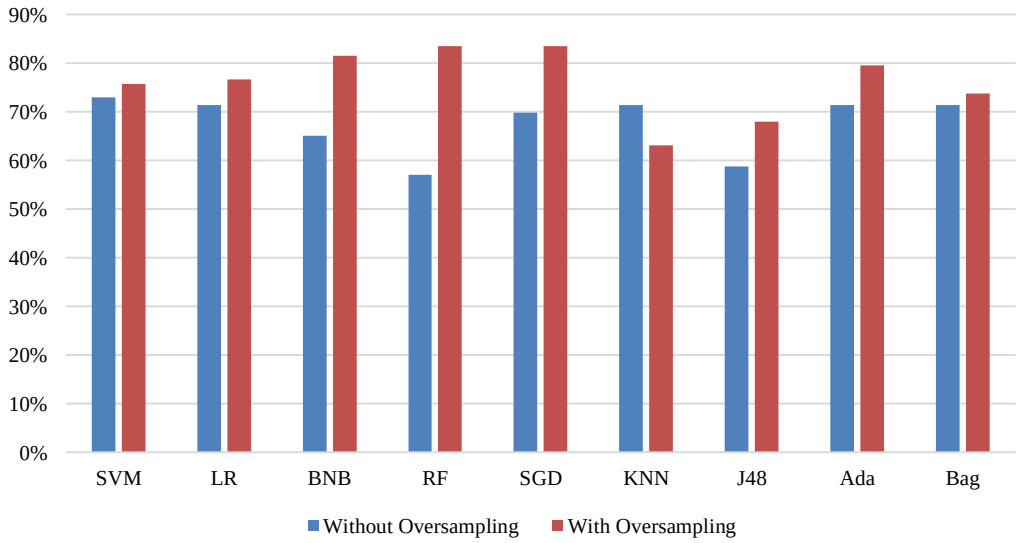

Fig. 6. Accuracy of Classifiers with 4-gram.

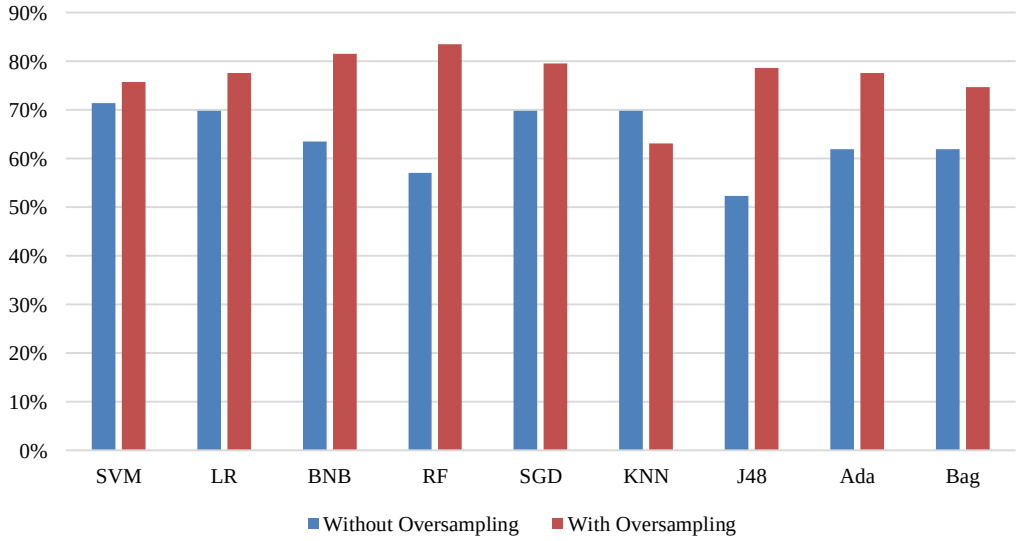

Fig. 7. Accuracy of Classifiers with 5-gram.

V. Conclusions and Future Works

This study investigated the performance of several machine learning methods to detect the health-related rumors in social media for Arabic language. The dataset (HRRD) was generated by extracting tweets regarding cancer disease from Twitter using Arabic language. The experiments were conducted by applying several preprocessing methods such as stemming, tokenization and oversampling. Then, several machine learning methods were applied. The experimental results showed that when the data is balanced (using oversampling method), the performance of machine learning methods clearly improved. The best accuracy was obtained by random forest classification (83.50%) using 4 and 5 gram as tokenization methods. Therefore, this study recommends using random forest to detect the health-related rumors in social media written in Arabic language. This study opens the door for other researchers to work on health-related rumors in Arabic and also provide the HRRD dataset available that can be also beneficial for further studies in health-related research. In future work, other machine learning methods can be applied with different preprocessing methods. In addition, the dataset can be enriched by including more tweets on cancer disease from social media.

References

[1] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36.

[2] Zhao, Y., & Zhang, J. (2017). Consumer health information seeking in social media: a literature review. Health Information & Libraries Journal, 34(4), 268-283.

[3] Kumar, S., & Shah, N. (2018). False information on web and social media: A survey. arXiv preprint arXiv:1804.08559.

[4] Kardos, P., Leidner, B., Pléh, C., Soltész, P., & Unoka, Z. (2017). Empathic people have more friends: Empathic abilities predict social

network size and position in social network predicts empathic efforts. Social Networks, 50, 1-5.

[5] Martínez-López, F. J., Anaya-Sánchez, R., Molinillo, S., Aguilar-Illescas, R., & Esteban-Millat, I. (2017). Consumer engagement in an online brand community. Electronic Commerce Research and Applications, 23, 24-37.

[6] Chiang, I. P., Wong, R., & Huang, C. H. (2019). Exploring the Benefits of Social Media Marketing for Brands and Communities. " International Journal of Electronic Commerce Studies", 10(2), 113-140.

[7] Utz, S., & Breuer, J. (2016). Informational benefits from social media use for professional purposes: Results from a longitudinal study. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 10(4).

[8] Nisar, T. M., Prabhakar, G., & Strakova, L. (2019). Social media information benefits, knowledge management and smart organizations. Journal of Business Research, 94, 264-272.

[9] Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., ... & Sandhu, M. (2019). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. Political Behavior, 1-23.

[10] Viviani, M., & Pasi, G. (2017). Credibility in social media: opinions, news, and health information—a survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7(5), e1209.

[11] Sharma, M., Yadav, K., Yadav, N., & Ferdinand, K. C. (2017). Zika virus pandemic—analysis of Facebook as a social media health information platform. American journal of infection control, 45(3), 301-302.

[12] Alhayan, F., Pennington, D. R., & Ayouni, S. (2018, April). Measuring passive engagement with health information on social media. In 2018 21st Saudi Computer Society National Computer Conference (NCC) (pp. 1-6). IEEE.

[13] Yang, M. (2019, July). Health information literacy of the older adults and their intention to share health rumors: an analysis from the perspective of socioemotional selectivity theory. In International Conference on Human-Computer Interaction (pp. 97-108). Springer, Cham.

[14] Gu, R., & Hong, Y. K. (2019). Addressing Health Misinformation Dissemination on Mobile Social Media.

[15] Trethewey, S. P. (2020). Strategies to combat medical misinformation on social media.

[16] Caulfield, T., Marcon, A. R., Murdoch, B., Brown, J. M., Perrault, S. T., Jarry, J., ... & Rachul, C. (2019). Health misinformation and the power of narrative messaging in the public sphere. Canadian Journal of Bioethics/Revue canadienne de bioéthique, 2(2), 52-60.

[17] Ma, J., Gao, W., Wei, Z., Lu, Y., & Wong, K. F. (2015, October). Detect rumors using time series of social context information on microblogging websites. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (pp. 1751-1754).

[18] Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. British journal of political science, 47(2), 241-262.

[19] H. Guo, J. Cao, Y. Zhang, J. Guo, and J. Li, ''Rumor detection with hierarchical social attention network,'' in Proc. 27th ACM Int. Conf. Inf. Knowl. Manage., 2018, pp. 943–951. doi: 10.1145/3269206.3271709.

[20] Thakur, H. K., Gupta, A., Bhardwaj, A., & Verma, D. (2018). Rumor detection on twitter using a supervised machine learning framework. International Journal of Information Retrieval Research (IJIRR), 8(3), 1-13.

[21] Al-Sarem, M., Boulila, W., Al-Harby, M., Qadir, J., & Alsaeedi, A. (2019). Deep Learning-Based Rumor Detection on Microblogging Platforms: A Systematic Review. IEEE Access, 7, 152788-152812.

[22] Alkhodair, S. A., Ding, S. H., Fung, B. C., & Liu, J. (2020). Detecting breaking news rumors of emerging topics in social media. Information Processing & Management, 57(2), 102018.

[23] Soon, J. J. Q., Banerjee, S., & Chua, A. Y. K. (2017). Analyzing medical personnel's perceptions of online health rumors.

[24] Chua, A. Y., & Banerjee, S. (2015, December). Analyzing users' trust for online health rumors. In International Conference on Asian Digital Libraries (pp. 33-38). Springer, Cham.

[25] Chua, A. Y., & Banerjee, S. (2018). Intentions to trust and share online health rumors: An experiment with medical professionals. Computers in Human Behavior, 87, 1-9.

[26] Zhou, J., Liu, F., & Zhou, H. (2018). Understanding health food messages on Twitter for health literacy promotion. Perspectives in public health, 138(3), 173-179.

[27] Kasemsap, K. (2017). Analyzing the role of health information technology in global health care. In Handbook of research on healthcare administration and management (pp. 287-307). IGI Global.

[28] Zhang, Z., Zhang, Z., & Li, H. (2015). Predictors of the authenticity of Internet health rumours. Health Information & Libraries Journal, 32(3), 195-205.

[29] Sicilia, R., Giudice, S. L., Pei, Y., Pechenizkiy, M., & Soda, P. (2018). Twitter rumour detection in the health domain. Expert Systems with Applications, 110, 33-40.

[30] Liu, Y., Yu, K., Wu, X., Qing, L., & Peng, Y. (2019). Analysis and Detection of Health-Related Misinformation on Chinese Social Media. IEEE Access, 7, 154480-154489.

[31] Armstrong, P. W., & Naylor, C. D. (2019). Counteracting health misinformation: a role for medical journals?. Jama, 321(19), 1863-1864.

[32] Li, Y., Zhang, X., & Wang, S. (2017). Fake vs. real health information in social media in China. Proceedings of the Association for Information Science and Technology, 54(1), 742-743.

[33] Zhang, Z., Zhang, Z., & Li, H. (2015). Predictors of the authenticity of Internet health rumours. Health Information & Libraries Journal, 32(3), 195-205.

[34] Daughton & Paul, 2019 Daughton, A. R., & Paul, M. J. (2019). Identifying Protective Health Behaviors on Twitter: Observational Study of Travel Advisories and Zika Virus. Journal of medical Internet research, 21(5), e13090.

[35] Ghenai & Mejova, Ghenai, A., & Mejova, Y. (2017). Catching Zika fever: Application of crowdsourcing and machine learning for tracking health misinformation on Twitter. arXiv preprint arXiv:1707.03778.

[36] Ghenai & Mejova, Ghenai, A., & Mejova, Y. (2018). Fake cures: user-centric modeling of health misinformation in social media. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 58.

[37] Mejova, 2018 Mejova, Y. (2018, April). Information Sources and Needs in the Obesity and Diabetes Twitter Discourse. In Proceedings of the 2018 International Conference on Digital Health (pp. 21-29). ACM.

[38] Alexander Kinsora, Kate Barron, Qiaozhu Mei, and VG Vinod Vydiswaran. 2017. Creating a Labeled Dataset for Medical Misinformation in Health Forums. In Healthcare Informatics (ICHI), 2017 IEEE International Conference on. IEEE, 456–461.

[39] Patty Kostkova, Vino Mano, Heidi J Larson, and William S Schulz. 2016. Vac medi+ board: Analysing vaccine rumours in news and social media. In Proceedings of the 6th International Conference on Digital Health Conference. ACM, 163–164.

[40] Samuel, H., & Zaïane, O. (2018). MedFact: Towards Improving Veracity of Medical Information in Social Media Using Applied Machine Learning. Lecture Notes in Computer Science, 108–120. doi:10.1007/978-3-319-89656-4_9.