Future Generation Computer Systems 117 (2021) 47–58

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

An ensemble machine learning approach through effective feature extraction to classify fake news

Saqib Hakak [a], Mamoun Alazab [b], Suleman Khan [c], Thippa Reddy Gadekallu [d,*], Praveen Kumar Reddy Maddikunta [d], Wazir Zada Khan [e]

[a] Canadian Institute for Cybersecurity, Faculty of Computer Science, University of New Brunswick, Fredericton, NB, Canada
[b] College of Engineering, IT & Environment, Charles Darwin University, NT, Australia
[c] Air University Islamabad, Islamabad 44000, Pakistan
[d] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India
[e] Jazan University, Department of CS & IS, Main University Campus, Jazan, SA 45142, Saudi Arabia

ARTICLE INFO

Article history:
Received 6 August 2020
Received in revised form 8 November 2020
Accepted 20 November 2020
Available online 24 November 2020

Keywords:
Fake news detection
Ensemble machine learning
Feature extraction
Liar dataset
ISOT dataset

ABSTRACT

There are numerous channels available such as social media, blogs, websites, etc., through which people can easily access the news. It is due to the availability of these platforms that the dissemination of fake news has become easier. Anyone using these platforms can create and share fake news content based on personal or professional motives. To address the issue of detecting fake news, numerous studies based on supervised and unsupervised learning methods have been proposed. However, all those studies do suffer from a certain limitation of poor accuracy. The reason for poor accuracy can be attributed due to several reasons such as the poor selection of features, inefficient tuning of parameters, imbalanced datasets, etc. In this article, we have proposed an ensemble classification model for detection of the fake news that has achieved a better accuracy compared to the state-of-the-art. The proposed model extracts important features from the fake news datasets, and the extracted features are then classified using the ensemble model comprising of three popular machine learning models namely, Decision Tree, Random Forest and Extra Tree Classifier. We achieved a training and testing accuracy of 99.8% and 44.15% respectively on the Liar dataset. For the ISOT dataset, we achieved the training and testing accuracy of 100%.

1. Introduction

The concept of fake news has been in existence even before the emergence of Internet and other computational technologies. Dissemination of fake news and misleading information has always been used as a weapon to fulfil immoral objectives since ages. The advancement of Internet and web technologies has made it very easy for anyone to post anything in online platforms like blogs, comments to news articles, social media, etc. The advancement of technologies has enabled convenient access to authentic and falsified information even faster posing a real challenge. The involvement of social media replacing the traditional media has an even more catalytic effect, where both fake and authentic news are spread extremely rapidly [1,2]. The spread of such fake news has extremely negative impact on target individuals and also the society at large. Consequently, it also creates an impression among readers such that the general perception and responses towards authentic news also gets diluted hampering the balance of news ecosystem [3–5]. One of the startling example is the US 2016 presidential election wherein fake news were purposely spread through Facebook and twitter at a larger scale in comparison to authentic information [6].

The spreading of such falsified information often act as a basis of deriving political strategies instead of traditional debates wherein civilized discussions lead to mutual agreements of facts. Such fake information affects stock prices, choice of stock purchases, investment plans and even reactions to natural calamities. The contents of fake news are framed to make believable information that would create mass opinions and completely convince the readers or make them utterly confused. Such information thus enforces readers to change their perceptions and reactions indirectly shifting attention from real news.

Extensive studies have been conducted to analyse the pattern of spreading fake information through social media as depicted in Fig. 1. But the need of empirical investigation towards its detection, identification of source and stopping of its spreading still

* Corresponding author.
E-mail addresses: saqib.hakak@unb.ca (S. Hakak), alazab.m@ieee.org (M. Alazab), 171518@students.au.edu.pk (S. Khan), thippareddy.g@vit.ac.in (T.R. Gadekallu), praveenkumarreddy@vit.ac.in (P.K.R. Maddikunta), wazirzadakhan@jazanu.edu.sa (W.Z. Khan).

https://doi.org/10.1016/j.future.2020.11.022
0167-739X/

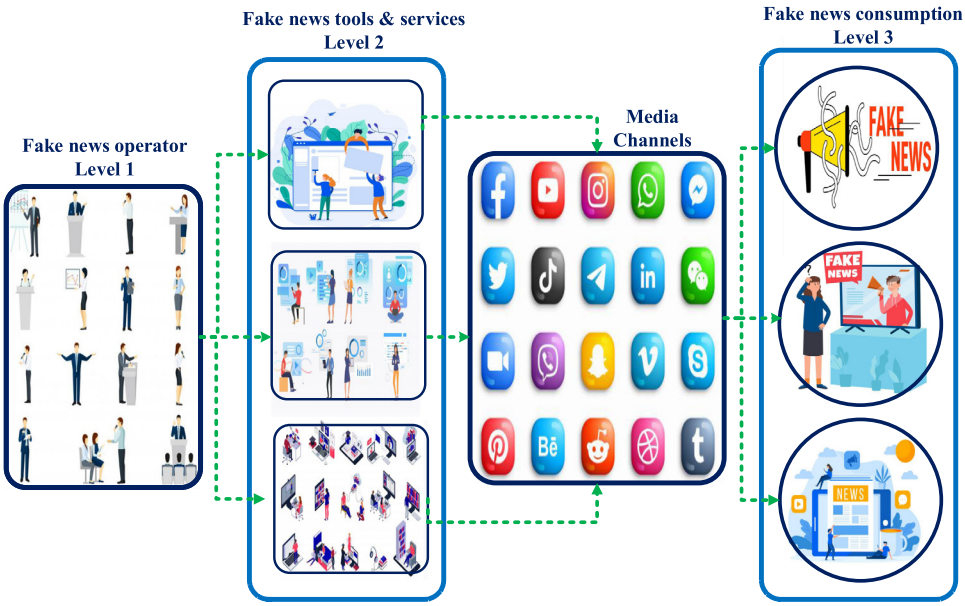


Fig. 1. Fake news levels.

exists. It is very important to stop circulation of misleading falsified information. This can primarily be achieved through human intervention by verifying authenticity of information using the International Fast Checking Network (ICFN) and other manual fast checking websites like Washington post, Snopes, Fast Checker, FastCheck and TruthOrFiction [7–9]. These websites are quite authentic and efficient, but they have scalability issues in handling large volumes of data. To overcome this aspect, the concept of automatic fact-checking has evolved consisting of three elements, namely, identification, verification and correction. These three elements work hand in hand to identify false claims, verify authenticity of the claims and delivers rectified genuine information across social media platforms. These automated fast checkers have agility in responding to information and are scalable in handling the high volume of news created across social media platforms [10].

Print media organizations also have developed their own customized web extensions like Decodex that help to segregate authentic and fake information. Fake news are often disseminated seeking help of automated algorithms and hence antidote algorithms also exist that help in identifying contents and sources of such information. These advanced algorithms are basically of three categories, content based, message diffusion dynamic based and hybrid. In spite the best efforts, all of these approaches have their limitations pertinent to absence of an all inclusive dataset having multidimensional information that would help to detect fake news characteristics with high accuracy. Machine learning approaches tend to yield higher accuracy and hence the present study investigates the authenticity of tweets validating the same with specific training features [3,11,12]. Strict enforcement of privacy and security [13–15] laws help in combating fake news in social media.

The present study involves experimentation on two popular fake news datasets, ISOT and Liar datasets, using Google Colab which is an open source cloud service provided by Google Inc. The data is cleaned and feature extraction is performed resulting in selection of the most significant features contributing towards detection of fake news. As part of the proposed algorithm, 70% of dataset is used for training and remaining 30% is used to test the classification model using k-fold cross validation. The extracted features are further classified using an ensemble machine learning model comprising of Decision Tree (DT) classifier, Random Forest (RF) algorithm and Extra Tree (ET) classifier. The main contributions of this paper are as follows:

1. Using feature extraction to use the most significant features that influence the classification of fake news.
2. Selection of an Ensemble model to achieve optimized accuracy in classification.
3. Reduction of training time of the ensemble classifier.

The organization of the paper is as follows - Section 2 provides extensive review of the relevant studies done in this topic of research. Section 3 presents the Background and proposed model followed by the Results and Discussion in Section 4 and the paper is concluded in Section 5.

2. Literature survey

The growth of social media is increasing exponentially, and most of the news is spread through social media instead of the standard media channels. Some of the news spread in social media is counterfeit, and this false information has a profound impact on the society, the government, political leaders, etc. Recently, several research works have been carried out using machine learning techniques to identify fake news.

The work in [16] introduces fake news detection in multiple languages, namely Slavic, Latin, and German, by selecting text features. The experiments were carried out on five different datasets, namely TwitterBR, FakeBrCorpus, FakeNewsData1, FakeOrRealNews, and btvlifestyle. Each dataset carried out different evaluations using custom features, Word2Vec, DCDistance, and bag-of-words. Finally, each dataset is fed to different classification algorithms such as KNN, Support Vector Machines (SVM), Random Forest (RF), and Gaussian Naive Bayes (NB). SVM and RF deliver better performance compared to conventional methods. However, the proposed model does not achieve a higher accuracy rate and thus restricts its use in the identification of fake news.

Another important research in [17] presents a new way of predicting fake news in social media well in advance. In this article, the authors proposed new classifier with different features, such as semantic features, user-based features, structural features, sentiment-based features, and predicted features. The

research was carried on an Italian Facebook dataset with 300 K official media news and 50 K posts from various blogs and websites providing false or incorrect facts. The results show the proposed early detection of fake news achieved 77% training accuracy and 91% testing accuracy. However, the fake news prediction can be enhanced by identifying the elements that negatively impact the information.

Another application of deep neural networks for early detection of fake news can be found in [18]. The status-sensitive crowd feedback acts as input during the process, and the Convolution Neural Network (CNN) is used as a classifier educated with positive and unlabelled samples. Five-fold cross-validation is used for model validation. The experiment was carried out on two different datasets, namely Twitter and Weibo. The proposed model achieved a 90% accuracy rate within 5 minutes of the news spread. However, the authors have used a small dataset which limited the proposed work.

More interestingly in [19], the authors proposed a supervised learning model for detecting fake news. The implementation suggests the addition of new features for training different classifiers that helps in identifying the fake news with a better accuracy rate. The experiment was carried out on 2282 news articles related to the US election. The results show that the RF classifier achieved an accuracy rate of 85%. However, the authors have used a relatively small dataset and have not applied supervised learning model on large datasets to evaluate the accuracy of deep learning approaches with improved fake news predictions.

Another framework for the detection of fake news using a deep learning approach can be found in [20]. The authors proposed Bidirectional Long Short Term Memory with Convolutional Neural Network (BiLSTM) to characterize tweets as fake news or not. The main objective of this model is to investigate tweets as fake or not by validating them with specific trained features. The authors used 5800 tweets during the investigation and the results show that the proposed model achieves 86.12% accuracy, 86%, recall and f-measure. However, we found that authors were limited to use only text-based features to classify the fake news. The selection of features can be enriched by using the image-based and video-based features. Artificial Intelligence (AI) and Neural Networks' tremendous growth has a significant impact on prediction models [21,22]. In [23], the authors proposed a Deep Structured Semantic Model (DSSM) with Recurrent Neural Networks (RNNs). The proposed model seeks to detect fake news by classifying news using semantic features. During this process, the authors used LIAR dataset, classifying the news using different classifiers. The results show that the proposed DSSM-LSTM model provides 99% training accuracy. However, the authors have not shown the feature extractions using semantic features, thus failing to use its validity in fake news predictions.

In another similar work in [24], the authors proposed a fake news prediction model focused on media news content, social media content, and the combination of both. The implementation is carried out on the BuzzFeed dataset, the PolitiFact dataset, where the input is fed to the 3-mode sensor, and the matrix-tensor factorization is used to extract the features. Furthermore, the extracted features are used for classification using XGBoost and DNN. The experimental results show that the proposed model achieves 85.86% accuracy on BuzzFeed dataset and 88.64% accuracy on PolitiFact dataset. The authors, however, did not demonstrate the feature extraction using content and context.

Another framework for fake news detection using machine learning techniques can be found in [25], where the authors used three approaches namely linguistics, sociocultural textual, and textual classification. In this process, the authors examined the binary classification using deep learning models, namely LSTM, RNN and Gated Recurrent Unit (GRU). Experimental findings show the proposed model achieves 75% training accuracy and 45% test accuracy for LSTM, 45% training accuracy and 42% test accuracy for GRU, 62% training accuracy and 48% test accuracy for RNN. However, we found that the authors failed to provide any information regarding automatically learning features.

In [26], the authors proposed a geometric deep learning model for the detection of false news. During this process, the authors used two convolutionary layers and two-dimensional features. The experiment was carried out on three different data sets, namely Buzzfeed, Snopes and PolitiFact, with the aim of classifying and detecting fake news at a higher accuracy rate. The results show that the proposed model achieved an accuracy rate of 92.7%. The proposed work can be extended in the future by classifying news based on topics and predicting fake news at an early stage.

The work in [27] introduces a neural network technique for the identification of fake news. The primary aim of this research is to identify fake news considering the length of the text and to apply different features, while non-static embedding is used during the learning phase. The research is carried out on two datasets, namely ISOT and liar datasets. The findings show that the proposed model achieves 99.1% accuracy for ISOT training dataset and 99.8% accuracy for ISOT testing dataset, 40.09% accuracy for liar training dataset and 39.5% accuracy for liar testing dataset.

Table 1 depicts the summary of existing fake news detection approaches.

3. Proposed methodology

In the proposed study, we identified twenty-six (26) linguistic-based textual features as listed in Table 2. The identified features were extracted from the text. For detecting fake news, the state-of-the-art machine learning models were explored. Fig. 2 presents the architecture of our proposed work. It consists of several phases such as data pre-processing, feature selection, ensemble-model selection, hyperparameter tuning and training of the model. Algorithms 1 and 2 presents the technical aspects of the proposed methodology. The brief description of the steps involved are as follows:

- Popular fake news datasets i.e. liar and ISOT were identified and explored for the proposed approach.
- To remove the noise from datasets, necessary preprocessing was done.
- After text preprocesing, tokenization is performed to convert the larger text in to words or in small lines.
- The major part of this research is extracting the features from text and then use these features for fake news detection instead of text.
- The extracted features are then passed to the state-of-the-art machine learning algorithms [28,29] like ensemble decision tree, random forest and extratree classifier to train the model.
- Various evaluation metrics are used to evaluate the performers of our proposed model.

3.1. Data cleaning

For this study, we explored two popular datasets i.e. Liar and ISOT. Liar dataset consists of labelled short statements covering various news topics that have been labelled manually. ISOT dataset is created by University of Victoria [30,31]. It contains 23,481 fake news articles and 21,417 true news articles respectively. True news articles have been taken from reliable sources such as reuters and fake news articles have been taken from the sites such as wikipedia and politifact that had flagged those articles.

Table 1
Summary of existing fake news detection approaches.

Reference	Dataset	Contributions	Accuracy Attained	Challenges
[16]	TwitterBR, FakeBrCorpus, FakeNewsData1, FakeOrReal News, and btv lifestyle	1. Fake news detection in multiple languages, 2. Each dataset carried out different evaluations using custom features, Word2Vec, DCDistance, bag-of-words and CNN is used as a classifier	79%	Does not attain higher accuracy
[17]	Italian Facebook dataset with 300K official media news, 50K incorrect information	Proposed a classifier with different features, such as semantic features, user-based features, structural features, sentiment-based features, and predicted features	91%	Failed to show identifying the element that negatively impact the information
[18]	Twitter and Weibo datasets	CNN is trained with positive and unlabelled samples. Five-fold cross-validation is used for model validation	90%	Limited to a tiny dataset of 1,111 Twitter posts and 816 Weibo posts
[19]	2282 news articles related to the US election	Addition of new features for training classifiers	85%	Limited to a tiny dataset. In the future, authors should be able to use a large number of datasets and apply deep learning approaches with improved fake news predictions
[20]	5800 twitter tweets	BiLSTM-CNN to characterize tweets as fake news or not by validating them with specific trained features	86.12%	limited to using only text-based features to classify fake news or not
[24]	BuzzFeed dataset, the PolitiFact dataset	Matrix-tensor factorization is used to extract the features. Furthermore, the extracted features are used for classification using XGBoost and DNN	85.86%	Did not demonstrate the feature extraction using content and context.
[23]	Liar dataset	DSSM-LSTM to detect fake news by classifying news using semantic features	99%	Failed to show feature extractions using semantic features
[25]	Extracted using FakeNewsNet tool	Examined the binary classification using deep learning models, namely LSTM, RNN, GRU	75%-LSTM 45%-GRU 62%-RNN	Does not attain higher accuracy, Limited to a tiny dataset
[26]	Buzzfeed, Snopes and PolitiFact	Geometric deep learning model with two convolutionary layers and two-dimensional features	92.7%	The proposed work can be extended in the future by classifying news based on topics and predicting fake news at an early stage
[27]	ISOT and LIAR	Neural network model, apply different feature extraction, non-static embedding is used during the learning phase.	99.8%-ISOT 39.5%-Liar	Limited to using only text-based features to classify fake news or not

Table 2
Features extracted from text.

Feature name	Data type	Feature name	Data type
Person	Numeric	NORP	Numeric
FAC	Numeric	Organization	Numeric
GPE	Numeric	Location	Numeric
Product	Numeric	Event	Numeric
Work of Art	Numeric	Law	Numeric
Language	Numeric	Date	Numeric
Time	Numeric	Percent	Numeric
Money	Numeric	Quantity	Numeric
Cardinal	Numeric	Ordinal	Numeric
word_count	Numeric	char_count	Numeric
sentence_count	Numeric	avg_word_length	Numeric
avg_sentence_length	Numeric	polarity	Numeric
avg_sentence_length	Numeric	sentiment_score	Numeric

3.1.1. Data preprocessing

In this phase, the given datasets were pre-processed to remove the noise such as stopwords, punctuation marks, html tags, url, emojis, etc. Pre-processing was done using NLTK toolkit which is an open-source and widely used NLP library. It comes with inbuilt functions and algorithms such as nltk.tokenize method (for tokenising text), nltk.stem.porter.PorterStemmer method (popular Porter stemming algorithm) and other such methods.

The pre-processing of datasets is carried out as follows:

- *Tokenization*: is the process of splitting the text/string into the list of tokens and is considered as the first step in natural language processing [32,33] before feature extraction process. nltk.tokenize method (an inbuilt function in nltk library) is used in this work for tokenization.
- *Stop Words Removal*: After tokenizing the text, this step consists of removing the stop words. Stop words are insignificant words in a language that create noise when used

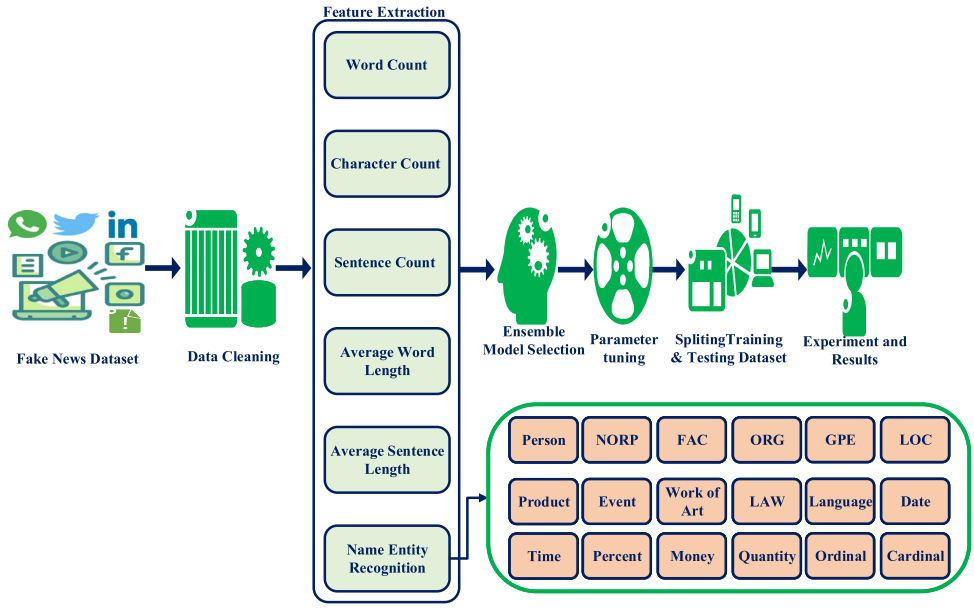


Fig. 2. Architecture of the proposed model.

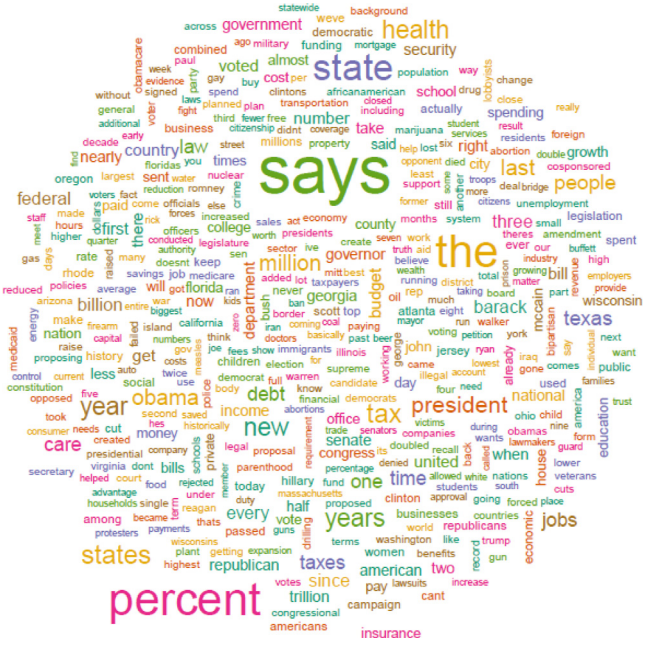


Fig. 3. Liar dataset real news word-cloud.

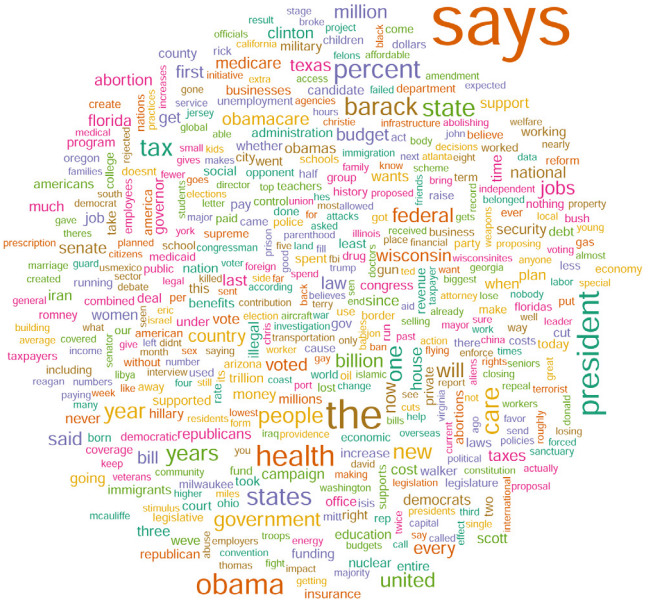


Fig. 4. Liar dataset fake news word-cloud.

as features in text classification. These words are frequently used in sentences to connect different words or to assist in the sentence structure. Articles, prepositions, and conjunctions and some pronouns are considered stop words. We removed common words such as a, about, an, are, as, at, be, by, for, from, how, in, is, of, on, or, that, the, these, this, too, was, what, when, where, who, will, and so on.

- *Stemming*: Stemming is a process of reducing the words to its root (also known as lemma). The main purpose of stemming is to reduce the frequency of derived words. For example, the words such as *running, ran, and runner* will be reduced to its lemma which is a word *run*. For this purpose, we used the porter stemmer algorithm, which is the most commonly used stemming algorithm.
- *Features Extraction*: Twenty-six features were identified for this study. The reason for selecting the less number of features was due to the fact that irrelevant features decreases the accuracy of the models and increases the cost of training process. Moreover selecting large number of features can also increase the training time of the models [34]. Hence, we selected less but effective features which includes: number of words, number of characters, number of sentences, average word length, average sentence length and Name Entity recognition-based features. For NER feature, we extracted the following information from the text: person, org, date, time, FAC (airports, buildings etc.), GPE (countries, cities etc.), product, work of art (book titles, song titles etc.), language, money and cardinal. Table 2 represents the number of NER features we extracted from the text.

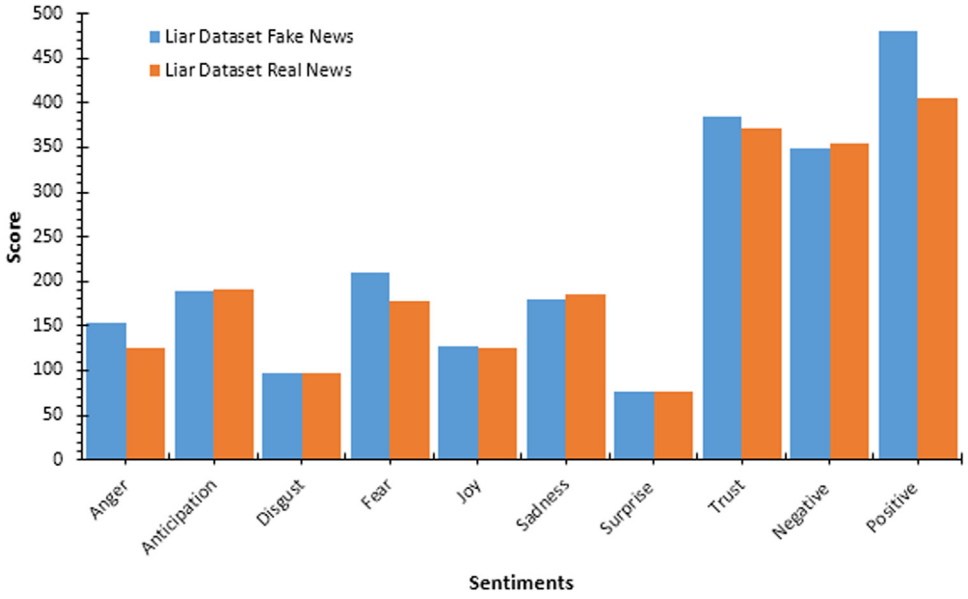


Fig. 5. Liar dataset fake and real news emotions.

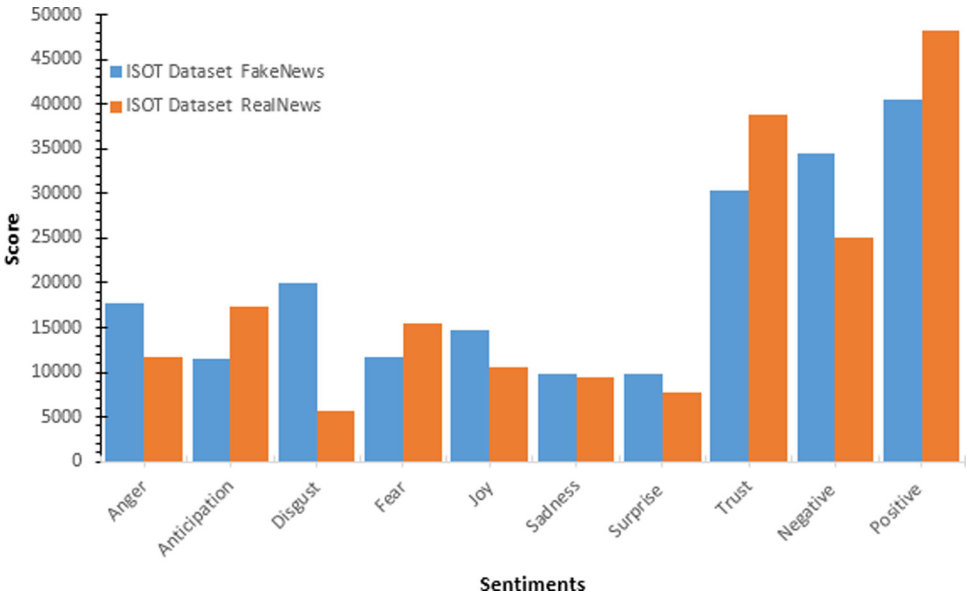


Fig. 6. ISOT Dataset fake and real news emotions.

3.2. Data visualization

To gain insight into the datasets, we carried out the visualization of datasets in the form word-cloud, pie charts, emotion graphs and frequency bar-graph. The purpose of visualization is to understand the structure of datasets. Fig. 3 and Fig. 4 represents word-cloud for fake news and real news for liar dataset respectively. Similarly Fig. 5 and Fig. 6 depict the sentiments of real and fake news for liar and ISOT datasets respectively. Fig. 7 represents the word-cloud for real news for ISOT dataset whereas Fig. 8 represents the word-cloud for fake news for ISOT dataset. Fig. 9 and Fig. 10 depict the class distribution of liar and ISOT datasets respectively.

3.3. Ensemble model selection

An ensemble approach is a technique that blends the predictions of several machine learning-based algorithms to make more accurate predictions [35–37]. In other words, we can say in an ensemble approach, multiple learning models are trained to create a single powerful predictive model. We identified and selected three popular supervised algorithms i.e. random forest, extra-tree algorithm and decision tree for the ensemble process. *Decision tree (DT) algorithm* is one of the popular algorithms used in supervised learning. It follows root to leaf representation and splits data continuously based on a certain parameter [38]. It comprises of roots, branches, leaves, etc. The leaf nodes are typically the class labels in a decision tree. [39] provides a further deep explanation of decision trees in supervised learning. On

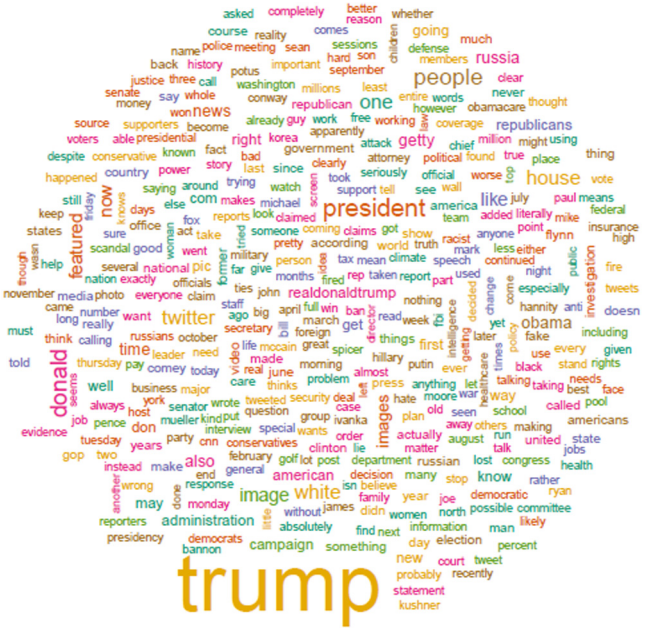


Fig. 7. ISOT Dataset real news word-cloud.

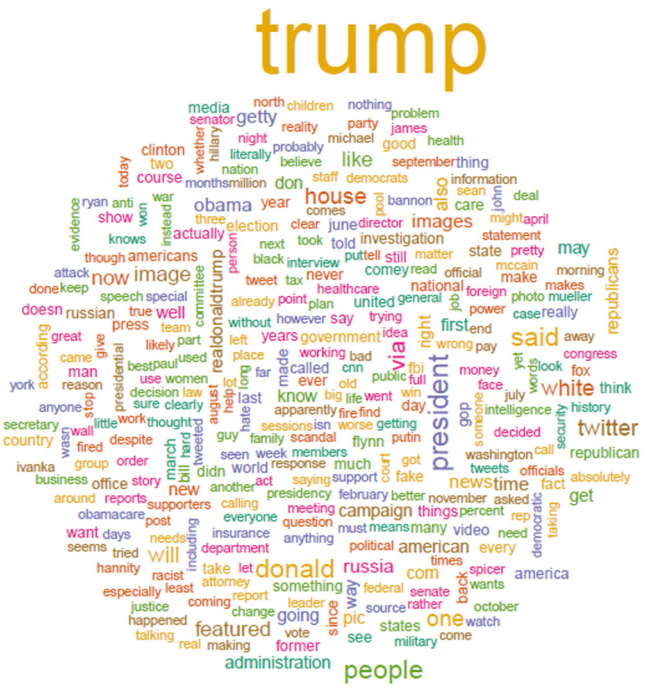


Fig. 8. ISOT dataset fake news word-cloud.

the other hand, the *Extra-Trees Algorithm (ETA)*, also known as Extremely Randomized Trees Classifier, constructs an unpruned decision ensemble or regression trees by traditional top-down method [40]. It aggregates the output into a forest from multiple de-correlated decision trees. At last, *Random Forest classifier* is a combination of multiple decision trees and provide more accurate outcomes [41]. To aggregate the output of these multiple models, a bagging approach was used. The motivation for using the bagging method is stability that it provides to the model and also it

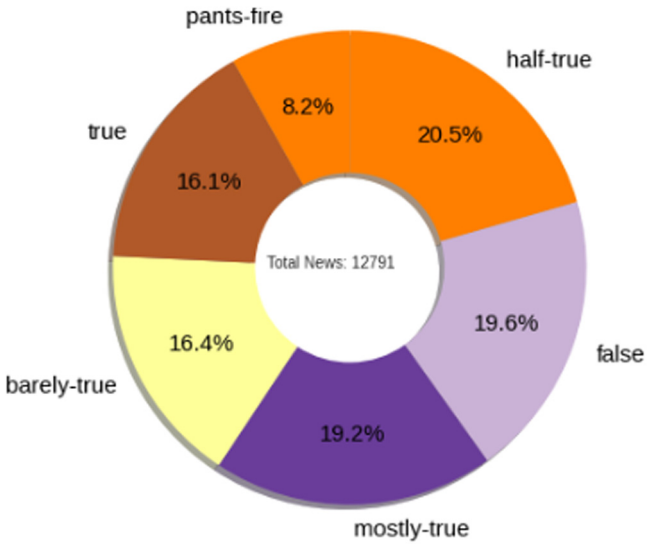


Fig. 9. Liar dataset class distribution.

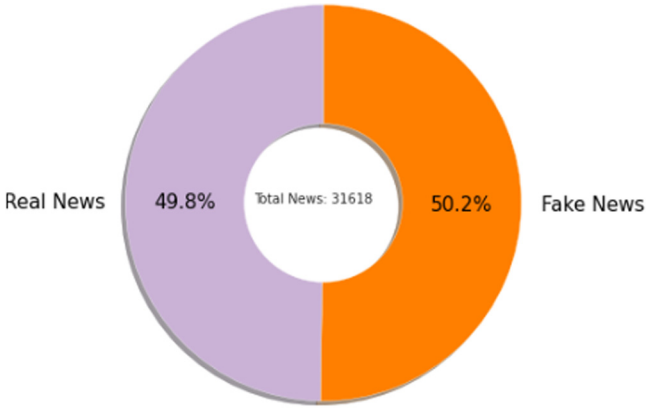


Fig. 10. ISOT Dataset class distribution.

reduces the possibility of overfitting in models. The algorithm for the proposed ensemble model is as follows:

$$\text{Entropy: } P(E_1, E_2, \ldots, E_n) = \sum_{i=1}^{n} E_i \log\left(\frac{1}{E_i}\right) \tag{1}$$

where P(E_1, E_2, E_3, ….., E_n) depicts the possible probability of target variable.

$$\text{Gain}(F, N) = P(F) \sum_{i=1}^{n} E(F_i)\, P(F_i) \tag{2}$$

3.4. Parameter tuning & training

Finally, after following all the above-mentioned steps, the datasets were divided into training set and testing set using the k-fold approach. Random search hyper-parameter tuning method is used for choosing optimal hyper-parameters. Hyper-parameters are the default values of the machine learning models that directly influence their performance during the training phase. Each model does come with the default values of parameters but that does not guarantee optimal performance. Also, it is unlikely to determine the optimal value of hyper-parameters before the training process. Therefore, different combinations were tested

Algorithm 1: Bagging Algorithm

Input: IOST and Liar datasets
Training:

1. Select Base classifier (Decision Tree, Random Forest and Extra Tree Classifier) and then select n number of samples for bagging.
2. Split the datasets in to two subsets (training and testing). Produce additional training datasets by using replacement sampling approach. These databases are $F_1, F_2, F_3, \ldots\ldots, F_n$. **Iteration**
3. Train a model classifier on a specified dataset F_i and construct n numbers of models $M_1, M_2, M_3, \ldots, M_n$. **return** *optimal Model*

Testing:

1. Every data entity X is transferred to trained models in the testing dataset, $M_1, M_2, M_3, \ldots, M_n$.
2. Every new data object is assigned with the label based on a majority vote. The majority vote is used to assign a new label to data point X for the classification problem and the average value is assigned to a new data object, X_i, which is used for the regression problem.
3. **while** *majority voting* **do**

We repeat those steps until every item is labelled in the dataset.

Table 3
Optimal hyperparamter.

Parameter	Proposed value
bootstrap	true
max_depth	100
max_features	auto
min_samples_leaf	5
min_samples_split	10
n_estimators	100
random_state	42
verbose	0
criterion	gini

for finding the optimal values. The optimal parameters chosen in the proposed work for several classifiers are depicted in Table 3.

For the evaluation purposes, accuracy, precision, recall and F-score performance metrics were considered. These metrics are briefly described as follows:

- *Accuracy* is the estimate of total number of correctly classified instances and is calculated using $(T_p + T_n)/(T_p + F_p + F_n + T_n)$.
- *Precision* is the percentage of relevant instances obtained from the total number of instances and is computed using $T_p/(T_p + F_p)$.
- *Recall* refers the percentage of relevant instances retrieved from the total number of relevant instances and computed using $T_p/(T_p + F_n)$.
- *F1 Measure* is the harmonic average of precision and recall given by:

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

Here T_p, F_n, F_p, T_n represent the number of true positives, false negatives, false positives and true negatives, respectively. The results are presented in the next section.

4. Experiment results

The experimentation on both ISOT and Liar datasets is performed in Google colab, a free GPU based cloud service offered by Google Inc. The programming language used for the experimentation is Python 3.7. In this work, for both the ISOT and Liar datasets, 70% of the records are used to train the classifiers and the remaining 30% of the records are used for testing the classifiers using k-fold cross validation. To classify the features extracted from the ISOT and Liar datasets, Decision Tree (DT) classifier, Random Forest algorithm and Extra Tree (ET) classifier are used. Rest of this section discusses about the results obtained on Liar and ISOT fake news datasets with and without feature extraction.

4.1. Experiment results for Liar dataset

In this sub-section, the performance evaluation of DT, RF, and ET classifiers on the Liar dataset with and without feature extraction are discussed.

The results obtained by implementing DT, RF, and ET classifiers on the Liar dataset (before and after extraction of the features) are depicted in Table 4.

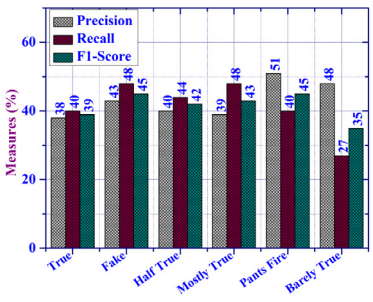
From Table 4 it can be observed that the Precision, Recall and F1-Score for DT classifier are 17%, 14%, and 16% for barely-true label, 22%, 27% and 24% respectively for false label, 24%, 21% and 22% respectively for half-true label, 22%, 24% and 23% respectively for mostly-true label, 20%, 15% and 17% respectively for pants-fire label, where as for the true label it is 20%, 22% and 21% respectively.

On implementing RF classifier on the Liar dataset without feature extraction are as follows: for the barely-true label, Precision, Recall and F1-score are 25%, 15%, 19% respectively, for the false label Precision, Recall and F1-score are 24%, 49%, 32% respectively, for the half-true label Precision, Recall and F1-score are 26%, 20%, 22% respectively, for the mostly-true label, Precision, Recall and F1-score are 27%, 31%, 29% respectively, for the pants-free label Precision, Recall and F1-score are 31%, 11%, 16% respectively, where as for the true label Precision, Recall and F1-score are 28%, 19%, 23% respectively.

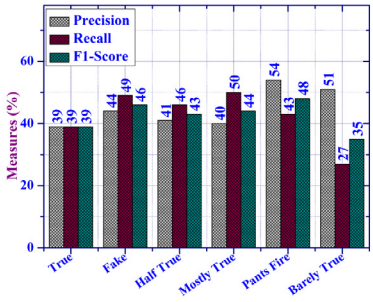
The experimental results of ET classifier on the Liar dataset without feature extraction are as follows: for the barely-true label, Precision, Recall and F1-score are 24%, 16%, 19% respectively, for the false label Precision, Recall and F1-score are 24%, 38%, 30% respectively, for the half-true label Precision, Recall and F1-score are 26%, 25%, 25% respectively, for the mostly-true label, Precision, Recall and F1-score are 26%, 29%, 27% respectively, for the pants-free label Precision, Recall and F1-score are 31%, 13%, 19% respectively, where as for the true label Precision, Recall and F1-score are 25%, 21%, 23% respectively.

From Table 4 it can be observed that the Precision, Recall and F1-Score for DT classifier after feature extraction are 38%, 39%, and 40% for barely-true label, 43%, 48% and 45% respectively for false label, 40%, 44% and 42% respectively for half-true label, 39%, 48% and 43% respectively for mostly-true label, 51%, 40% and 45% respectively for pants-fire label, where as for the true label it is 48%, 27% and 35% respectively.

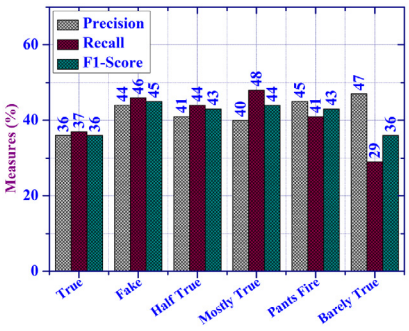
The obtained results on implementing RF classifier on the Liar dataset after feature extraction are as follows: for the barely-true label, Precision, Recall and F1-score are 39%, 39%, 39% respectively, for the false label Precision, Recall and F1-score are 44%, 49%, 46% respectively, for the half-true label Precision, Recall and F1-score are 41%, 46%, 43% respectively, for the mostly-true label, Precision, Recall and F1-score are 40%, 50%, 44% respectively, for the pants-free label Precision, Recall and F1-score are 54%, 43%, 48% respectively, where as for the true label Precision, Recall and F1-score are 41%, 27%, 35% respectively.



(a) Classification Report for Bagging Decision Tree



(b) Classification Report for Random Forest



(c) Classification Report for Extra Tree Classifier

Fig. 11. Classification report of Liar dataset for different classifiers.

Table 4
Classification report for liar dataset.

Classifier	Label	Before features extraction			After features extraction		
		Precision	Recall	F1-Measure	Precision	Recall	F1-Measure
Decision tree	barely-true	17	14	16	38	39	40
	false	22	27	24	43	48	45
	half-true	24	21	22	40	44	42
	mostly-true	22	24	23	39	48	43
	pants-fire	20	15	17	51	40	45
	true	20	22	21	48	27	35
Random forest	barely-true	25	15	19	39	39	39
	false	24	49	32	44	49	46
	half-true	26	20	22	41	46	43
	mostly-true	27	29	31	40	50	44
	pants-fire	31	11	16	54	43	48
	true	28	19	23	41	27	35
Extra tree classifier	barely-true	24	16	19	36	37	36
	false	24	38	30	44	46	45
	half-true	26	25	25	41	44	43
	mostly-true	26	29	27	40	48	44
	pants-fire	31	13	19	45	41	43
	true	25	19	23	47	29	36

The experimental results of ET classifier on the Liar dataset after feature extraction are as follows: for the barely-true label, Precision, Recall and F1-score are 36%, 37%, 36% respectively, for the false label Precision, Recall and F1-score are 44%, 46%, 45% respectively, for the half-true label Precision, Recall and F1-score are 41%, 44%, 43% respectively, for the mostly-true label, Precision, Recall and F1-score are 40%, 48%, 44% respectively, for the pants-free label Precision, Recall and F1-score are 45%, 41%, 43% respectively, where as for the true label Precision, Recall and F1-score are 47%, 29%, 36% respectively.

The accuracy and the training time of DT, RF, and ET classifiers for the Liar dataset, before and after feature extraction are depicted in Table 5.

From the table, it can be observed that before feature extraction, RF and ET classifiers achieve better prediction accuracy than DT classifier. Concerning the prediction time, the ET classifier outperforms the other two classifiers considered. The training accuracy of these classifier is 99.93% each. Where as the ET classifier outperforms the other two classifiers in terms of the training time of 2.5 s.

Table 5
Accuracy and the training time for liar dataset.

Classifier	Accuracy and time complexity before feature extraction				Accuracy and time complexity after feature extraction			
	Prediction accuracy	Prediction time (s)	Training accuracy	Training time (s)	Prediction accuracy	Prediction time (s)	Training accuracy	Training time (s)
Decision tree	21.23	1.26	99.93	15.94	42.15	0.01	99.96	0.099
Random forest	25.92	1.42	99.93	25.66	44.15	0.40	99.96	1.70
Extra tree classifier	25.20	0.02	99.93	2.58	42.20	0.52	99.96	1.59

Table 6
Classification report for ISOT Dataset.

Classifier	Label	Before features extraction			After features extraction		
		Precision	Recall	F1-Measure	Precision	Recall	F1-Measure
Decision tree	Fake	100	99	100	100	100	100
	Real	99	100	99	100	100	100
Random forest	Fake	99	97	98	100	100	100
	Real	97	99	98	100	100	100
Extra tree classifier	Fake	97	98	98	100	100	100
	Real	97	98	98	100	100	100

Table 7
Accuracy and the training time for ISOT dataset.

Classifier	Accuracy and time complexity before feature extraction				Accuracy and time complexity after feature extraction			
	Prediction accuracy	Prediction time (s)	Training accuracy	Training time (s)	Prediction accuracy	Prediction time (s)	Training accuracy	Training time (s)
Decision tree	99.29	0.05	100	5.36	100	0.01	100	0.32
Random forest	98.45	4.04	100	36.58	100	0.60	100	4.67
Extra tree classifier	97.59	4.78	100	65.09	100	0.01	100	0.32

Table 8
Comparison of the proposed work with existing works.

Ref.	Method	Dataset		ISOT Dataset		Liar dataset	
				Accuracy			
		ISOT Dataset	Liar Dataset	Training accuracy	Testing accuracy	Training accuracy	Testing accuracy
[23]	Deep Structured Semantic Model (DSSM) with Recurrent Neural Networks(RNNs)	X	✓	X	X	99%	X
[42]	IntentCapsNet	X	✓	X	X	X	24.6%
[43]	LSTM model	X	✓	X	X	X	38.5%
[27]	Capsule neural networks	✓	✓	99.1%	99.8%	X%	40.9%
Proposed	Ensemble machine learning with feature extraction	✓	✓	100%	100%	99.96%	44.15%

It can be observed from the table that after feature extraction, the accuracy of all the classifiers has been increased, and also the training time for all the classifiers has been reduced. RF has better accuracy and when compared with the other two classifiers whereas DT has less training time.

4.2. Experiment results for ISOT dataset

In this sub-section the performance evaluation of DT, RF, and ET classifiers on ISOT dataset with and without feature extraction are discussed. Table 6.

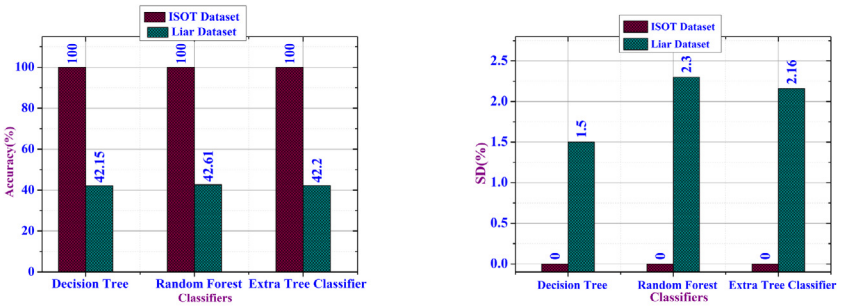
It is evident from the table that DT classifier outperforms the other two classifiers in all the three measures, Precision, Recall and F1-Score with 100%, 99% and 100% for the fake news and 99%, 100% and 99% for the real news.

The prediction accuracy, training accuracy, prediction time, and training time are presented in Table 7. From the table, it can be observed that before feature extraction, the DT classifier achieves a better prediction accuracy of 99.29% when compared with RF and DT which achieve a prediction accuracy of 98.45% and 97.59% respectively. The training accuracy of all the three classifiers is 100%. DT outperforms the other two classifiers in terms of prediction time and training time. After feature extraction, all the classifiers achieve 100% prediction and training accuracy. In terms of prediction and training times for the ISOT dataset after feature extraction, DT and ET classifiers perform better with 0.32 s.

Fig. 11 presents a comparative analysis of all the labels for the three classifiers on the Liar dataset. Fig. 12 summarizes the accuracy and Standard Deviation (SD) achieved by all the three classifiers for both the Liar and ISOT datasets.

4.3. Comparative analysis of proposed model with existing works

Table 8 presents the comparative analysis of the present work with existing works. From the table it can be observed that the current work outperforms the existing works on both the datasets considered, ISOT dataset and Liar dataset. The proposed method achieves a training and testing accuracy of 100% on ISOT dataset and 99.6% training accuracy and 44.15% testing accuracy on Liar dataset.



(a) Accuracy (b) SD

Fig. 12. Comparative analysis of accuracy and standard deviation for Liar and ISOT datasets.

4.4. Discussion

Feature extraction from the text represents the text messages which simplifies the classification of the text data. The feature extraction process reduces the dimensionality of the text thereby eliminating irrelevant features from the text data. This aids in improving the accuracy of the classifiers and also the reduction of training time. In this work, to improve the classification accuracy of machine learning algorithms and also to reduce the training time for fake news datasets, number of words, number of characters, number of sentences, average word length, average sentence length, and name entity recognition-based features are extracted from ISOT and Liar datasets. The experiment results prove that by extracting important features from fake news datasets, the performance of the machine learning algorithms has been improved.

5. Conclusion

In this article, we proposed a machine-learning based fake news detection model using a supervised approach. We used the ensemble approach for training and testing purposes consisting of decision tree, random forest, and extra tree classifiers. The aggregation of outputs was done using the bagging approach and compared to the state-of-the-art, our model achieved better results. Precise feature selection and hyperparameter tuning were the two important phases that contributed in achieving better performance. The experimentation of the model using liar and ISOT datasets yielded an accuracy of 44.15% and 100% percent. In the future, we aim to extend the proposed work by adding more datasets and also using hyperparameter tuning models based on meta-heuristic algorithms.

CRediT authorship contribution statement

Saqib Hakak: Software, Writing - original draft. **Mamoun Alazab:** Conceptualization, Methodology, Software. **Suleman Khan:** Data curation, Writing - original draft. **Thippa Reddy Gadekallu:** Visualization, Investigation, Supervision. **Praveen Kumar Reddy Maddikunta:** Writing - review & editing. **Wazir Zada Khan:** Software, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Buntain, J. Golbeck, Automatically identifying fake news in popular Twitter threads, in: 2017 IEEE International Conference on Smart Cloud, SmartCloud, IEEE, 2017, pp. 208–215.
[2] B. Liu, J.D. Fraustino, Y. Jin, Social media use during disasters: A nationally representative field experiment, Tech. Rep., College Park, MD, 2013.
[3] N.X. Nyow, H.N. Chua, Detecting fake news with tweets' properties, in: 2019 IEEE Conference on Application, Information and Network Security, AINS, IEEE, 2019, pp. 24–29.
[4] S.I. Manzoor, J. Singla, et al., Fake news detection using machine learning approaches: A systematic review, in: 2019 3rd International Conference on Trends in Electronics and Informatics, ICOEI, IEEE, 2019, pp. 230–234.
[5] Q. Chen, G. Srivastava, R.M. Parizi, M. Aloqaily, I. Al Ridhawi, An incentive-aware blockchain-based solution for internet of fake media things, Inf. Process. Manage. 57 (6) (2020) 102370.
[6] A. Bovet, H.A. Makse, Influence of fake news in Twitter during the 2016 US presidential election, Nat. Commun. 10 (1) (2019) 1–14.
[7] G. Kessler, About the Fact Checker, Vol. 10, The Washington Post, 2011.
[8] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, J. Econ. Perspect. 31 (2) (2017) 211–236.
[9] G. Gravanis, A. Vakali, K. Diamantaras, P. Karadais, Behind the cues: A benchmarking study for fake news detection, Expert Syst. Appl. 128 (2019) 201–213.
[10] D. Funke, Here's how Close Automated Fact-Checking is to Reality, Poynter, 2018.
[11] S.M.S. Shah, T.A. Malik, R. Khatoon, S.S. Hassan, F.A. Shah, Human behavior classification using geometrical features of skeleton and support vector machines, Comput. Mater. Continua 61 (2) (2019) 535–553.
[12] M.B. Nejad, M.E. Shiri, A new enhanced learning approach to automatic image classification based on Salp Swarm Algorithm, Comput. Syst. Sci. Eng. 34 (2) (2019) 91–100.
[13] M. Alazab, R. Layton, R. Broadhurst, B. Bouhours, Malicious spam emails developments and authorship attribution, in: 2013 Fourth Cybercrime and Trustworthy Computing Workshop, 2013, pp. 58–68.
[14] M. Alazab, S. Huda, J. Abawajy, R. Islam, J. Yearwood, S. Venkatraman, R. Broadhurst, A hybrid wrapper-filter approach for malware detection, J. Netw. 9 (11) (2014) 2878–2891.
[15] K.-N. Tran, M. Alazab, R. Broadhurst, et al., Towards a Feature Rich Model for Predicting Spam Emails Containing Malicious Attachments and Urls, Australian Computer Society Inc., 2014.
[16] P.H.A. Faustini, T.F. Covões, Fake news detection in multiple platforms and languages, Expert Syst. Appl. (2020) 113503.
[17] M.D. Vicario, W. Quattrociocchi, A. Scala, F. Zollo, Polarization and fake news: Early warning of potential misinformation targets, ACM Trans. Web (TWEB) 13 (2) (2019) 1–22.
[18] Y. Liu, Y.-F.B. Wu, FNED: A deep network for fake news early detection on social media, ACM Trans. Inf. Syst. (TOIS) 38 (3) (2020) 1–33.
[19] J.C. Reis, A. Correia, F. Murai, A. Veloso, F. Benevenuto, Supervised learning for fake news detection, IEEE Intell. Syst. 34 (2) (2019) 76–81.
[20] M.Z. Asghar, A. Habib, A. Habib, A. Khan, R. Ali, A. Khattak, Exploring deep neural networks for rumor detection, J. Ambient Intell. Humaniz. Comput. (2019) 1–19.
[21] G. Raja, Y. Manaswini, G.D. Vivekanandan, H. Sampath, K. Dev, A.K. Bashir, AI-powered blockchain-a decentralized secure multiparty computation protocol for IoV, in: IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS, IEEE, 2020, pp. 865–870.
[22] S.S. Zehra, R. Qureshi, K. Dev, S. Shahid, N.A. Bhatti, Comparative analysis of bio-inspired algorithms for underwater wireless sensor networks, Wirel. Pers. Commun. (2020).

[23] S.S. Jadhav, S.D. Thepade, Fake news identification and classification using DSSM and improved recurrent neural network classifier, Appl. Artif. Intell. 33 (12) (2019) 1058–1068.
[24] R.K. Kaliyar, A. Goswami, P. Narang, DeepFakE: improving fake news detection using tensor decomposition-based deep neural network, J. Supercomput. (2020).
[25] A. Vereshchaka, S. Cosimini, W. Dong, Analyzing and distinguishing fake and real news to mitigate the problem of disinformation, Comput. Math. Organ. Theory (2020) 1–15.
[26] F. Monti, F. Frasca, D. Eynard, D. Mannion, M.M. Bronstein, Fake news detection on social media using geometric deep learning, 2019, arXiv preprint arXiv:1902.06673.
[27] M.H. Goldani, S. Momtazi, R. Safabakhsh, Detecting fake news with capsule neural networks, 2020, arXiv preprint arXiv:2002.01030.
[28] T.R. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P.K. Reddy Maddikunta, I.-H. Ra, M. Alazab, Early detection of diabetic retinopathy using PCA-firefly based deep learning model, Electronics 9 (2) (2020) 274.
[29] T. Reddy, S.P. RM, M. Parimala, C.L. Chowdhary, S. Hakak, W.Z. Khan, et al., A deep neural networks based model for uninterrupted marine environment monitoring, Comput. Commun. (2020).
[30] H. Ahmed, I. Traore, S. Saad, Detection of online fake news using n-gram analysis and machine learning techniques, in: International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, Springer, 2017, pp. 127–138.
[31] H. Ahmed, I. Traore, S. Saad, Detecting opinion spams and fake news using text classification, Secur. Privacy 1 (1) (2018) e9.
[32] P.K.R. Maddikunta, T.R. Gadekallu, A. Al-Ahmari, M.H. Abidi, et al., Location based business recommendation using spatial demand, Sustainability (2071-1050) 12 (10) (2020).
[33] M.Z. Asghar, F. Subhan, H. Ahmad, W.Z. Khan, S. Hakak, T.R. Gadekallu, M. Alazab, Senti-eSystem: A sentiment-based eSystem-using hybridized fuzzy and deep neural network for measuring customer satisfaction, Softw. - Pract. Exp. (2020).
[34] G.T. Reddy, M.P.K. Reddy, K. Lakshmanna, R. Kaluri, D.S. Rajput, G. Srivastava, T. Baker, Analysis of dimensionality reduction techniques on big data, IEEE Access 8 (2020) 54776–54788.
[35] G.S. Aujla, R. Chaudhary, N. Kumar, R. Kumar, J.J. Rodrigues, An ensembled scheme for QoS-aware traffic flow management in software defined networks, in: 2018 IEEE International Conference on Communications, ICC, IEEE, 2018, pp. 1–7.
[36] G.T. Reddy, S. Bhattacharya, S.S. Ramakrishnan, C.L. Chowdhary, S. Hakak, R. Kaluri, M.P.K. Reddy, An ensemble based machine learning model for diabetic retinopathy classification, in: 2020 International Conference on Emerging Trends in Information Technology and Engineering, Ic-ETITE, IEEE, 2020, pp. 1–6.
[37] C. Iwendi, S. Khan, J.H. Anajemba, A.K. Bashir, F. Noor, Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model, IEEE Access 8 (2020) 28462–28474.
[38] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, S. Mishra, Decision tree and SVM-based data analytics for theft detection in smart grid, IEEE Trans. Ind. Inf. 12 (3) (2016) 1005–1016.
[39] T.G. Dietterich, et al., Ensemble learning, in: The Handbook of Brain Theory and Neural Networks, Vol. 2, MIT Press Cambridge, Massachusetts, 2002, pp. 110–125.
[40] X. Dong, Z. Yu, W. Cao, Y. Shi, Q. Ma, A survey on ensemble learning, Front. Comput. Sci. (2020) 1–18.
[41] G. Biau, E. Scornet, A random forest guided tour, Test 25 (2) (2016) 197–227.
[42] C. Xia, C. Zhang, X. Yan, Y. Chang, P. Yu, Zero-shot user intent detection via capsule neural networks, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 3090–3099, http://dx.doi.org/10.18653/v1/D18-1348, URL https://www.aclweb.org/anthology/D18-1348.
[43] Y. Long, Q. Lu, R. Xiang, M. Li, C.-R. Huang, Fake news detection through multi-perspective speaker profiles, in: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Vol. 2: Short Papers), Asian Federation of Natural Language Processing, Taipei, Taiwan, 2017, pp. 252–256, URL https://www.aclweb.org/anthology/I17-2043.

SAQIB HAKAK received the bachelor's degree in computer science engineering from the University of Kashmir, India, in 2010, and the master's degree in computer and information engineering from IIUM, Malaysia, and the Ph.D. degree from the Faculty of Computer Science and Information Technology, University of Malaya, Malaysia. He is currently working as an Assistant Professor with the University of Northern British Columbia, Canada. Prior to this designation, he worked as a Postdoctoral Research Fellow with the prestigious Canadian Institute for Cyber-Security. His research interests include information security, natural language processing, cyber security, artificial intelligence, and wireless networks.

MAMOUN ALAZAB is Associate Professor at the College of Engineering, IT and Environment at Charles Darwin University, Australia. He received his Ph.D. degree in Computer Science from the Federation University of Australia, School of Science, Information Technology and Engineering. He is a cyber security researcher and practitioner with industry and academic experience. Alazab's research is multidisciplinary that focuses on cyber security and digital forensics of computer systems with a focus on cybercrime detection and prevention including cyber terrorism and cyber warfare. He has more than 100 research papers. He delivered many invited and keynote speeches, 22 events in 2018 alone. He convened and chaired more than 50 conferences and workshops. He works closely with government and industry on many projects, including Northern Territory (NT) Department of Information and Corporate Services, IBM, Trend Micro, the Australian Federal Police (AFP), the Australian Communications and Media Authority (ACMA), Westpac, United Nations Office on Drugs and Crime (UNODC), and the Attorney General's Department. He is a Senior Member of the IEEE. He is the Founder and Chair of the IEEE Northern Territory (NT) Subsection.

SULEMAN KHAN received the master's degree from the Department of Computer Science, Air University, Islamabad, Pakistan, in 2019. He is currently a Research Associate with Air University. His research interests include network security, machine learning, and data science.

THIPPA REDDY GADEKALLU received the B.Tech. degree in CSE from Nagarjuna University, the M.Tech. degree in CSE from Anna University, Chennai, India, and the Ph.D. degree from VIT, Vellore, India. He is currently working as an Associate Professor with the School of Information Technology and Engineering, VIT. He has 14 years of experience in teaching. He has authored/co-authored more than 50 international/national publications. His current areas of research are machine learning, deep neural networks, the Internet of Things, and blockchain.

M. PRAVEEN KUMAR REDDY received the B.Tech. degree in CSE from JNT University, and the M.Tech. degree in C.S.E. and the Ph.D. degree from VIT, Vellore, India. He was a Visiting Professor with the Guangdong University of Technology, China, in 2019. He had worked as a Software Developer with IBM, in 2011. He worked with Alcatel-Lucent, in 2013. He produced more than 15 international/national publications. He is currently working as an Assistant Professor with the School of Information Technology and Engineering, VIT. He is also working in the area of energy aware applications for Internet of Things (IoT) and high-performance computing.

WAZIR ZADA KHAN (Senior Member, IEEE) received the bachelor's and master's degrees in computer science from COMSATS University Islamabad, Wah Campus, in 2004 and 2007, respectively, and the Ph.D. degree from the Electrical and Electronic Engineering Department, UniversitiTeknologiPETRONAS, Malaysia, in 2015. He is currently working with the Farasan Networking Research Laboratory, Faculty of CS and IT, Jazan University, Saudi Arabia. His current research interests include wireless sensor networks, security and privacy, and the IoT.