Journal of Autonomous Intelligence (2024) Volume 7 Issue 3
doi: 10.32629/jai.v7i3.798

ORIGINAL RESEARCH ARTICLE

Automatic detection of fake news using recurrent neural network—Long short-term memory

Anil Kumari Shalini[1], Sameer Saxena[1], Billakurthi Suresh Kumar[2,*]

[1] *Department of Information Technology, Amity Institute of Information Technology, Amity University, Rajasthan, Jaipur 302006, India*

[2] *Department of Computer Science and Engineering, School of Computer Science and Engineering, Sanjay Ghodawat University, Kolhapur 416118, India*

*** Corresponding author:** Billakurthi Suresh Kumar**,** sureshkumarbillakurthi@gmail.com

ABSTRACT

The propagation of deliberate misinformation is gaining significant momentum, especially across different social media platforms. Despite the fact that there are several fact-checking blogs and websites that distinguish between news that is genuine or fake, and regardless of the reality that there is ongoing research being conducted to restrict the propagation of fake news, the issue is still one that needs to be addressed. The most major barrier is a failure to spot monitors and to disclose false news in a reasonable timeframe, both of which are critical components. In this research, a system is proposed that utilizes deep learning model of Recurrent Neural Network—Long Short-Term Memory (RNN-LSTM) in order to put an end to the circulation of misleading information and put a stop to, and expose instances of fake news that are spread through reliable channels. The process of extracting hybrid features from text data, such as Lemmas, Bi-Gram, Tri-Gram, N-gram, Term Frequency Inverse Document Frequency (TF-IDF), part-of-speech, and dependency-based natural language processing features, is developed as a strategy. When compared to other traditional approaches to machine learning classifier, the RNN-LSTM method that was proposed obtains a greater level of accuracy than those other approaches. It achieves an accuracy rate of 99.10% both during training and testing, which is better to the accuracy achieved by common machine learning approaches such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), Naïve Bayes (NB), and Decision Tree (DT). In the proposed approach for detecting false news using RNN-LSTM, the three experiments are conducted to acquire, accuracy, precision, recall, and F-score with varying forms of cross validation (5-Fold, 10-Fold and 15-Fold). Based on the findings of the empirical research, the conclusion can be drawn that the RNN-LSTM with ReLu function provides more accurate detection than both the RNN-LSTM (Tan h) function and the RNN-LSTM (sigmoid) function with 15 fold cross validation.

Keywords: fake news detection; machine learning classifiers; RNN-LSTM

ARTICLE INFO

Received: 7 July 2023
Accepted: 4 September 2023
Available online: 27 December 2023

1. Introduction

The introduction of the World Wide Web and the swift proliferation of platforms for social media (like Instagram and Twitter) prepared the way for a level of information transmission that had never occurred before in the records of human history. In addition to other use cases, news publishers benefited from the broad adoption of social media sites since it enabled them to provide their customers with updated news in a manner that was nearly identical to real time. The traditional print forms of the news media, like journals, tabloids, and newspapers, have given way to their digital counterparts, which include online news portals, websites, twitter feeds, as well as other digital content formats[1]. Consumers now have a simpler time getting

their hands on the most recent news at their fingers. Seventy percent of all traffic to news sites comes from referrals on Facebook. It needs to identify either it may fake or genuine news using supervised machine learning techniques[2]. These social media sites in its current iterations, are incredibly strong and beneficial due to their capacity to enable users to discuss and exchange ideas, as well as engage in debate over topics such as democracy, schooling, and healthcare[3]. People's ability to make decisions is largely dependent on the kinds of information they take in; likewise, their perspectives on the world are formed by the kinds of information they digest. There is growing evidence that people have reacted irrationally to news that was later proven to be fake[4,5]. This data comes from a number of different sources.

Furthermore, such channels are also utilized in a deceptive manner by certain organizations, most frequently for the purpose of financial gain. In other instances, such platforms are used to create prejudiced viewpoints, manipulate mindsets, and propagate satire or ridiculousness. The phenomenon is most frequently referred to as fake news[6]. In the past decade, there has been a fast increase in the propagation of fake news, which was most vividly noticed during the elections that took place in the United States in 2016. This widespread dissemination of internet publications that do not correspond to the facts has resulted in a number of issues that are not confined to the realm of politics but instead extend to a variety of other fields, including athletics, medicine, and science[7]. The stock markets are one sector that is susceptible to the influence of fake news. In this sector, a rumour can have catastrophic effects and may even force a halt to trading activity.

The good news is that there are a variety of computational methods that may be used to identify particular articles as being fraudulent based on the textual content of those articles[8]. The vast majority of these strategies make use of fact-checking sites like "PolitiFact" and "Snopes." There are a variety of archives that are maintained on a regular basis by researchers, and some of these repositories include records of websites that have been determined to be questionable or fraudulent. However, the difficulty with these resources is that in order to determine whether articles or websites are fraudulent, human skill is required. Even more crucially, the websites that check facts only contain articles from specific domains, such as politics, and they are not generalized to identify false news stories from other domains, such as amusement, games, and technology.

Documents, movies, and audio recordings are just some of the different types of media that may be found on the World Wide Web. It is relatively challenging to locate and classify content that has been published on the internet in an unstructured manner, as this task absolutely necessitates the involvement of human expertise. However, computational methods like Natural Language Processing (NLP) can be utilized to identify anomalies that distinguish text articles that are inherently dishonest from those that are founded on truth. Analysis of the dissemination of fake news in comparison to the actual news is another technique[9]. A strategy that is more hybrid can also be performed to examine the social response to a post along with analyzing the textual elements to determine whether or not an article is misleading in its nature. This can be done by hybrid approach[10]. Another recent example of this is the propagation of the new corona virus, which was accompanied by the dissemination of false reports across the Internet regarding the origin, existence, and behaviour of the virus. As more readers share about the bogus contents online, the situation became even more precarious[11]. Finding news like this on the internet can be a difficult process.

To be more precise, the method investigates the ways in which a piece of fake news spreads faster on a network in comparison to a piece of genuine news[12]. On a conceptual level, the response that a piece of writing receives can be differentiated to determine whether or not the writing in question is genuine.

The subject of detecting false news is addressed and a solution to this issue is proposed in this research by using RNN-LSTM approach. In this study, a number of different textual features that have the potential to be exploited to differentiate between actual and fraudulent content are investigated. With the help of those characteristics, a number of distinct Deep Learning (DL) and Machine Learning (ML) algorithms are trained with the assistance of a number of different ensemble approaches. Because learning models have a propensity

to lower error rate by employing strategies such as boosting and bagging[13], ensemble learners have been shown to be beneficial in a broad variety of applications. This is due to the fact that the ensemble learners are able to train in groups[14]. These methods make it possible to train a variety of machine learning techniques in a way that is both successful and efficient. On datasets that are accessible to the public, a comprehensive range of experiments are carried out. The findings provide evidence that the proposed method offers an increased performance when measured against the four-performance[15] metrics that are most frequently applied: precision, F1-score, recall, and accuracy.

The structure of this paper is fragmented into 5 parts. In part 2, various existing work of fake new detection using machine and deep learning technique is reviewed. In part 3, the framework of proposed RNN-LSTM model for detection of fake news is discussed in detail. In part 4, experimental findings and discussion of the proposed model is illustrated. Finally in part 5, conclusion of the proposed work is discussed.

2. Literature survey

Using the LIAR dataset, Wynne and Swe[1] developed a system that may be used to construct an ensemble classification method for the purpose of identifying fake news with a higher level of accuracy. The findings of the experiments indicate that the two-layered ensemble classifier that was developed achieves a level of accuracy that is 2.3% higher than that achieved by previous efforts.

Using the decision tree machine learning technique, Krishna and Adimoolam[2] find out exact Fake News Identification on social media. In addition, textual property correctness is compared with the support vector machine ML techniques. As part of the suggested research, an investigation into the detection of fake news was carried out using various machine learning techniques, including the DT algorithm (N = 311) and the SVM classifier (N = 311). The accuracy measures for the DT method appear to be 97.67%, which is a higher percentage than the accuracy measures for the SVM algorithm, which appear to be 91.74%. There is a difference between the research groups that may be considered significant, with a statistical significance of 92% for accuracy and 82.5% for precision for a confidence interval of 95%. Therefore, the SVM method is not utilised in the process of determining whether or not a piece of news obtained from social networking site is phoney. Instead, the DT method is used because its accuracy appears to be superior to that of the SVM classifier.

According to the findings of Mladenova and Valova[3], the capacity to recognize the distinction between real and fabricated information is a skill that develops naturally with age and life experience. Children are instructed from a young age on how to recognize lies and other forms of dishonesty. In addition, for many years, people have had the ability to identify lying in general. Why should people be waiting a whole day for the copy of the newspaper perhaps several hours for the nightly news when they can check into Instagram and see from their feed? Media, magazines, and television are the last sources of data in this day and age. After all, this is the reason why people have registered to the news source's social media page in the first place. A survey with 137 participants of varying ages and origins, such as students in high school college students, and graduate students, who worked in a variety of occupations such as technicians, financial advisors, teachers, specialists, designers, etc., was carried out. The participants came from a wide range of educational and occupational backgrounds. The purpose of this study is to investigate whether or not individuals who use social media have the ability to recognize clickbait and fake news. An investigation of this nature might prove valuable in the creation of machine learning algorithms for the classification of fake news, awareness raising, and tools for the identification and prevention of fake news. The level of data literacy amongst general public could be determined by carrying out a study on a wider scale that is analogous to the one that is now being carried out.

Renuka and Anithaashri[4] improves the identification of fake news in social networking sites by utilizing approaches from artificial intelligence, and their results are compared to those of the apriori algorithm. The

effectiveness analysis was carried out using a sample size of (N = 10), and it was evaluated with apriori (N = 10). The findings were evaluated based on the precision of each algorithm. Both support vector machine methods and apriori algorithms have significance levels that have values (*p* less than 0.05) that result in greater performance. The SVM method, which has a detection accuracy of 91.87%, is superior to the apriori algorithm, which has a detection accuracy of 31.76%, when it comes to detecting fake information in social media. The execution of this project demonstrates an improved detection using the SVM algorithm which is much greater than that of the apriori algorithm.

Gupta et al.[5] compared and evaluated the efficiency of a variety of machine learning techniques by using a variety of performance criteria. The research investigates a number of linguistic characteristics that, when combined, can be used to distinguish between fake and genuine news. Natural Language Processing methods are utilized during the pre-processing of data, which ultimately results in an increase in the precision of machine learning models. In addition, the extracted and precompiled attributes are used to train a variety of machine learning classifiers with all of the possible combinations, and the constructed models are therefore evaluated utilizing a variety of performance criteria.

Tian and Baskiyar[6] offer a technique for the detection of fake news that makes use of a K-nearest Neighbours (KNN) machine algorithms. The maximum level of accuracy that could be reached in this research was 91.3% when the algorithm for detecting fake news utilized Genetic and Evolutionary Feature Selection, also known as GEFeS. In addition, it has trained and tested a quantum KNN (QKNN) using the GEFeS discovered characteristics and an ideal *k* value in order to investigate how quantum ML methods may be applied to the challenge of detecting fake news. The QKNN model has a success rate of 84.4% in terms of accuracy.

Agarwal et al.[7] carried out an initial examination into the capabilities of five different ML models by using the LIAR dataset. In addition, the effectiveness of two feature extraction strategies, namely the Count Vectorizer as well as the Term Frequency-Inverse Document Frequency (TF-IDF) Vectorizer, was evaluated. The findings revealed that the TF-IDF Vectorizer performed marginally better than the Count Vectorizer.

Chaturvedi et al.[8] describe a ground-breaking approach to the detection of fake news that makes use of machine learning techniques. According to the findings of the studies, the method under consideration makes use of the algorithm known as the passive aggressive method in order to correctly classify the results. This kind of algorithm is always passive, but when there is an error in computation, it is becoming active and begins to update and adjust. The objective here is to make adjustments that will rectify the issue while having only a little impact on the final results. This will put us in a better position to attain the desired results. The investigations and tests that were carried out indicate that this solution has an accuracy of approximately 96%.

The trust notion for social networks and trust-related difficulties with prediction processes are covered by Choudhary et al.[9]. Author categorizes the trust forecast by resolving the problems and requests contributions for this age. The Zervopoulos et al.[10] new technologies develop daily, a need for viral reduction approach is identified in order to manage the deception of society by fake news. Shrivastava et al. develop a methodology to assess the spread of fake news and explain how it affects various categories. The present epidemic is referred to the authors as COVID-19 for viral false news.

Using Logistic Regression (LR), Krishna[16] accomplish effective Fake News Detection and evaluate textual feature accuracy using the SVM technique. In the research that was proposed, the analysis for detecting fake news was carried out utilizing machine learning techniques including the LR method (N = 311) and the SVM algorithm (N = 311) with G power 80% and alpha value 0.05. The LR and SVM classifiers were used to conduct an investigation on the reliability of fake news. The LR method appears to have an accuracy of 95.12%, whereas the SVM algorithm seems to have an accuracy of 91.68%. There is a statistical significance value

among the sample groups, with a significant level of 0.079 for correctness and 0.125 for precision. This indicates that there is a difference. In light of this, it would appear that the LR method is more effective than the SVM classifier when it comes to determining whether or not the news is fake.

Mahmud et al.[17] give a comparison analysis between many widely utilized machine learning methods and Graph Neural Networks for the purpose of identifying the propagation of misleading information on social media sites. For the context of this research, the UPFD dataset was utilized, and numerous previously developed machine learning algorithms were applied to text data only. In addition to this, a variety of GNN layers were developed in order to fuse graph-structured media dispersion data with text data in order to use the text data as that of the node attribute in GNN models. In the context of the research, GNNs offer the most effective answers to the challenge of recognizing fake news.

The purpose of the study by Billones et al.[18] is to reduce the amount of people who believe false news by investigating the viability of utilizing NB and Stochastic gradient classifier models to determine if an article written in English or Filipino is authentic or not. This is performed by training the models using big datasets that have already been pre-processed. Following the completion of the evaluation, it was determined that both models had attained an accuracy of 93% and 95%, correspondingly.

In recent years, Vinothkumar et al.[19] have found that online news has supplanted traditional news outlets as the most important source of information. People today don't have the time to go through the newspaper, so they turn to social media in order to be informed about what's going on in the world. On the other hand, information found on the internet can be ambiguous at times, and it may even be designed to mislead users. In the current system, the utilization of automated technologies for the detection of false news, like models that utilize machine learning, has become obligatory. The effectiveness of ML models was assessed using hold out cross validation on two datasets of varied sizes, one including fake news and the other containing actual news. The innovative stacking model that was suggested achieved testing accuracy of 99.9% on the ISOT dataset and 96% on the KD nugget dataset, correspondingly. It was not possible to produce an accurate result when using the dataset to identify false propaganda from current affairs, and the method was only capable of recognizing fake news. However, the system was able to identify fake news. Regarding the particular group, monitoring the tweets in real time allowed for the detection of fake news. The global method is capable of gathering data about broad feelings and is applied to a number of tweets simultaneously. Greedy and Dynamic Blocking Algorithms, all of which are exclusive to Trends, as well as models based on Support Vector Machines, were used. Additionally, sentiment knowledge was gathered from both unlabelled and labelled specimens within each Trend, and it utilizes this data to enhance the learning of sentiment categorization that is specific to Patterns. In this strategy for facilitating the interchange of sentiment data across important key words, we make advantage of restoration performed over Trends-specific classification models for determining how people feel about things.

The purpose of the paper by Divadkar et al.[20] is to conduct a study of the existing models for identifying ambiguous or fake news by utilizing deep learning, ML, and ensemble learning methodologies. In this review, a large number of contributions were compared using a few important factors, and investigated the issues that is presented. According to their analytical observations, the majority of the works employed the Kaggle dataset for the execution of their ideas. The accuracy results of machine learning-based educational environments and ensemble based learning models were surpassed by the results of deep learning based systems.

Simone et al.[21] utilises a Passive-Aggressive-Classifier, a RF, and an LSTM network, all of which have been trained to differentiate between fake and non-fake (actual) news. In addition, these models are utilised to categorise news sources in terms of the potential amount of false information that they could disseminate. The models are validated using German articles that have been translated into English. The Passive-Aggressive-Classifier yields the greatest results when used to articles written in English while attempting to detect false

news. Random Forest produces the greatest results when applied to the task of automatically ranking German-translated news articles. RF's accuracy in predicting actual news rankings attained a correlation of 0.68. This demonstrates that the application of automatic categorization may be adapted to work with languages other than English by applying this methodology. In the not too distant future, this method will be expanded upon by utilizing many different types of machine learning and translators.

According to Gupta et al.[22], fake news is a piece of data that purposefully spreads misleading information. The inspiration can come from anything, from political propaganda to the pursuit of personal gain. Because of this, a solution has been developed that is based on Deep Learning, which is a massive upgrade from the typical ML technique. By utilizing these deep learning techniques, it was possible to determine whether or not a piece of information was accurate. In order to accomplish this goal, Convolutional Neural Network (CNN) and LSTM were utilized as the basis algorithms. Ensembling was utilized as a methodology via the soft voting method. Ensembling is a model that has its correctness, which increases the total probability of accuracy for the dataset that was given. In addition to the base methods that were employed, this method has been used. The model is validated using a sizable database from Kaggle, and it is trained using data from a variety of news websites. This is the latest development in the approach known as machine learning.

A false news identification strategy based on headlines and news body is discussed by Umer et al.[23]. For improved results, the author additionally took into consideration the dimensional reduction strategy in addition to using principal component analysis and chi-square to extract quality characteristics. PCA is used to remove noise and talks about model gain with a 97.8% accuracy rate.

Han and Mehta[24] analyze and assess the effectiveness of machine learning and deep learning algorithms in identifying bogus news in social networks. Due to the quickest and most convenient means of information transmission, false news spreads quickly across society and influences people's opinions. The reader's brain is significantly affected by the false information because of the manipulative elements. Naive Bayes, a hybrid convocational neural network, and a recurrent neural network technique are used by the authors.

Malicious social bots are disseminating false information to deceive society, according to Sahoo SR et al.[25], hence it is important to identify and get rid of these bots from social networks. Typically, this detection employs quantitative behavioural analysis characteristics that are easily copied and has low detection accuracy. Using transition probability-based feature selection and semi supervised clustering for detection, the author describes the combined strategy.

3. Research methodology

The framework of proposed RNN-LSTM model for detecting fake news is illustrated in **Figure 1**. Initially the data is collected from the LIAR dataset which is publicly available on Kaggle. This data is pre-processed to remove noise, duplicate and redundant data using techniques namely stop word removal, tokenization, stemming etc. The pre-processed data is then splitted into 70:30 ratio, where 70% of data is used during training phase and 30 percent of the data is used during testing phase. The splitted data is given as an input to feature extraction and selection phase were variety features of dataset are extracted. Finally in the classification phase, news is classified into either fake news or normal news. The detail description of each stage of proposed framework is discussed in detail as follows.

Data collection: The LIAR dataset is utilised for the proposed model. This dataset consists of three files in.csv format: one each for testing, training, and validating the model. This dataset was obtained through the application programming interface (API) of the website that checks facts, PolitiFact. It contains 12,836 human brief comments that have been tagged. These statements are taken from a variety of contexts, such as press releases, interviews on television or radio public statements, and so on. The degrees of honesty in the media are categorised into a number of subcategories, including false, slightly true, half true, mainly true, and true.

Data pre-processing: The data collected from social media platforms is generally unstructured, and the vast bulk of it consists of informal conversation that contains things like typos, slang, and syntax errors. The pursuit of higher levels of performance and dependability has made it absolutely necessary to devise methods for optimizing resource use in order to arrive at well-informed decisions. Before the data can be utilized for predictive modelling, it is required to clean it up in order to improve the insights that can be gained from it. In order to achieve this goal, the News training data underwent some essential preprocessing. The cleansing of the data has been the emphasis of this step.

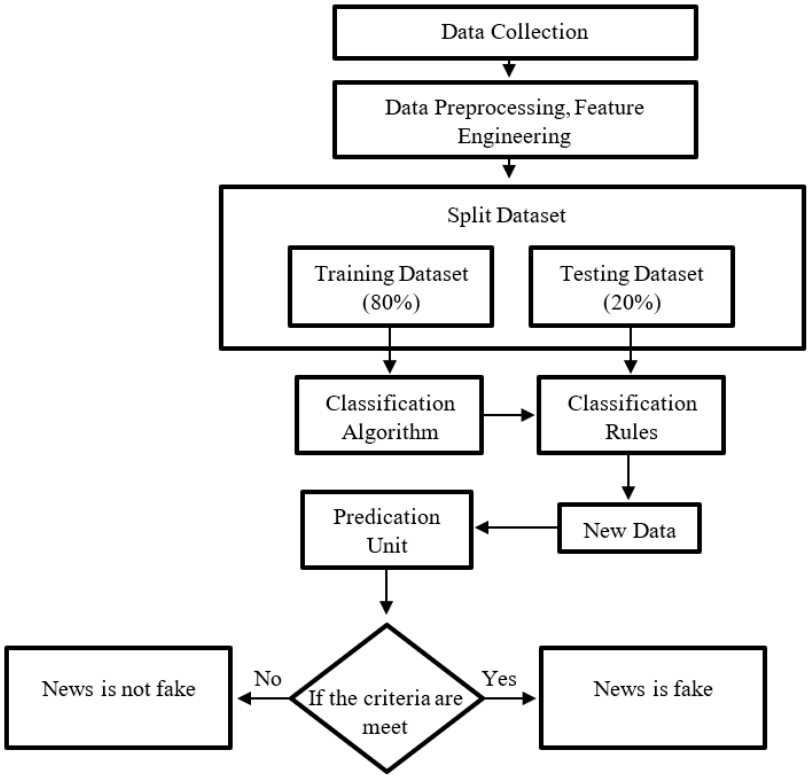

Figure 1. Framework of proposed RNN-LSTM model for detecting fake news.

Data cleaning: Data can either be organised in a structured fashion or in an unstructured fashion. A structured format follows a pattern that is clearly established, whereas unstructured material does not follow any particular pattern. When compared to an unstructured manner, a semi-structured format provides comparable levels of structure. It is required to clean up the textual information in order to emphasise the attributes. The process of data cleaning, also known as preprocessing, often involves the following steps[25]:

Remove punctuation: A sentence's punctuation can also provide grammatical context that strengthens the reader's understanding of the statement. However, because the vectorizer just counts the amount of words instead of the context, it does not contribute value; hence, every special characters are stripped away e.g., where are you. After removing punctuation the generated statement will be like this where are you?

Tokenization: The process of tokenizing a text divide it up into individual pieces such as statements or words. It provides structure to text that was not previously structured.

Remove stopwords: Common terms that are expected to exist in any text are referred to as stopwords. It does not provide us with very much information on the data, therefore it is removed. e.g.: gold or diamond is good for me after removing stopwords from above statement we got the words such as gold, diamond, good.

Stemming: The process of stemming significantly minimize a word to its root form, known as its stem. It is common sense to handle words that are linked in the same manner. Through the application of a

straightforward rule-based system, it eliminates suffices such as 'ed', 'ing,' 'ly,' and so on. The number of words in the corpus is decreased, but the actual words are frequently ignored. eg: borrowing, borrowed and after stemming algorithm it becomes a borrow.

Feature extraction and selection: The creation of a hybrid feature set is the ultimate outcome of the process of extracting features from a dataset. During this process, a number of different features are taken from the dataset that is being utilized for training. The following features were collected from the data:

Term frequency based features: The meanings of phrases that recur multiple times in a body of text may provide crucial insights into the context. When designing many different types of systems, one common method for selecting characteristics is to utilise the frequency of a phrase as a parameter.

N-gram features: Text mining and NLP are two applications that requires major use of this feature in their processes. N-grams are simply groups of words that have the same frequency of occurrence in the same window as one another. These are utilised in a wide variety of settings and applications. These ideas can be applied not just to the construction of unigram models, but also to the construction of bigram models and trigram models.

Bi-tagged features: In order to get started with the process of extracting bi-tagged qualities, the first thing is to tag phrases with a Part of Speech (POS) tagger. The components of a phrase that are referred to as "bitagged" are those that demonstrate a significant relationship between two words that follow each other in the phrase in a sequential order.

POS tag-based features: The POS tagging of textual material is essential to the operation of a wide variety of applications for NLP, overview generation, and request to response. Words that are made up of nouns, adjectives, adverbs, and verbs are effective at conveying data to the reader more about context, particularly information that is relevant to the reader. Therefore, text mining methods can considerably benefit from having such key features.

Classification: Proposed RNN-LSTM model and various machine learning/deep learning classifiers are used to classify the news into fake news or normal news. The performance of the model is evaluated using performance measurements namely accuracy, precision, f-score and recall.

Therefore, the below Algorithm 1 is proposed.

Algorithm 1 RNN-LSTM hybrid classifiers algorithm for classification text data

1: **Input:** Normalized training dataset *Train_Data[],* Normalized testing dataset *Test_Data[],* defined threshold *qTh*

2: **Output:** Result set as output with *{Predicted_class, weight}*

3: **Step 1:** Read all test data from *Test_Data[]* using below function for validating to training rules, the data is normalized and transformed according to algorithms requirements

4: $$test\_Feature(data) = \sum_{m=1}^{n} \quad (Attribute\_Set[A[m] \ldots\ldots A[n] \quad Test\_Data)$$

5: **Step 2:** select the features from extracted attributes set of $test\_Feature(data)$ and generate feature map using below function.

6: $$\text{Test\_FeatureMap } [t\ldots\ldots n] = \sum_{x=1}^{n} \quad (t) \; \square \; test\_Feature(x)$$

7: $Test\_FeatureMap\ [x]$ are the selected features in pooling layer. The convolutional layer extracts the features from input and passes to pooling layer and those selected features are stored in *Test_FeatureMap*

8: **Step 3:** Now read entire taring dataset to build the hidden layer for classification of entire test data in sense layer,

9: $$train\_Feature(data) = \sum_{m=1}^{n} \quad (Attribute\_Set[A[m] \ldots\ldots A[n] \quad Train\_Data)$$

10: **Step 4:** Generate the training map using below function from input dataset

Algorithm 1 (*Continued*)

11: Train_FeatureMap [t…..…n] = $\sum_{x=1}^{n}$ (t) □ $train\_Feature$(x)

12: $Train\_FeatureMap[t]$ is the hidden layer map that generates feature vector for build the hidden layer. That evaluate the entire test instances with train data.

13: **Step 5:** After generating the feature map we calculate similarity weight for all instances in dense layer between selected features in pooling layer

14: $$Gen\_weight = CalcWeight\ (Test\_FeatureMap\ ||\ \sum_{i=1}^{n}\ Train\_FeatureMap[i])$$

15: **Step 6:** Evaluate the current weight with desired threshold

16: $$if\,(Gen\_weight >= qTh)$$

17: **Step 7:** $Out\_List.add$ **(**$trainF.class, weight$**)**

18: **Step 8:** Go to step 1 and continue when Test_Data == $null$

19: **Step 9:** Return Out_List

4. Result and discussion

The implementation has done utilising open source java environment. For the goals of this study, a distributed Intel i3 CPU with 2.8 gigahertz of processing speed and 4 gigabytes of random-access memory was utilised. For the purpose of carrying out an experiment, 546 different news items are currently being dissected in order to be categorised. The method that was proposed was only successful in correctly categorising 500 of a total of 546 items of news. As a result of this, the proposed approach has a detection rate of 99.1%.

Consider the below **Table 1**, **Figure 2** which illustrates the performance measurements of various conventional Machine learning and Deep Learning techniques and proposed RNN-LSTM classification model.

Table 1. Performance analysis of proposed RNN-LSTM and other ML/DL techniques.

Algorithms	Accuracy	Precision	Recall	F1-score
Random forest	0.969	0.975	0.981	0.972
Decision tree	0.975	0.986	0.983	0.976
Artificial neural network	0.969	0.983	0.971	0.973
Support vector machine	0.976	0.986	0.985	0.989
Naive bayes	0.962	0.981	0.979	0.982
DNN	0.986	0.987	0.989	0.99
PNN	0.972	0.975	0.981	0.982
RNN-LSTM	0.991	0.987	0.989	0.990

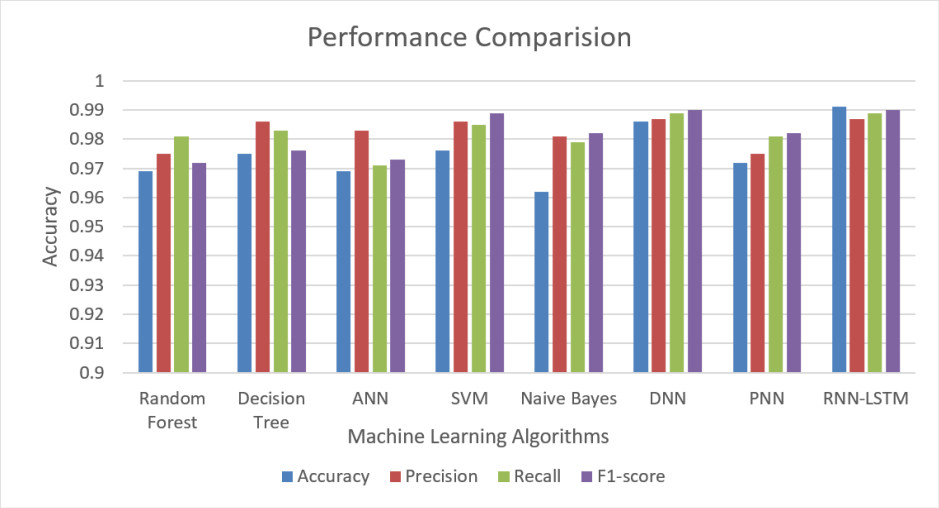

Figure 2. Performance analysis of proposed RNN-LSTM and other ML/DL techniques.

Table 1 and **Figure 2** show that the performance of proposed RNN-LSTM classifier for detecting fake news is 99.1% which is better as compared to standard classifiers namely random forest, ANN, naïve bayes, decision tree and support vector machine, deep neural network and probabilistic neural network etc.

In the proposed approach for detecting false news using RNN-LSTM, the three experiments are conducted to acquire, accuracy, precision, recall, and f-score with varying forms of cross validation, which are discussed as follows.

4.1. Experiment with the RNN-LSTM algorithm (sigmoid)

In this experimentation of the RNN-LSTM (sigmoid) framework, accuracy, precision, recall, and f-score are achieved with various forms of cross validation. Consider the validation of the model shown in **Figure 3**, which was carried out with 5, 10, and 15 folds of cross validation with the RNN-LSTM (sigmoid) classification model.

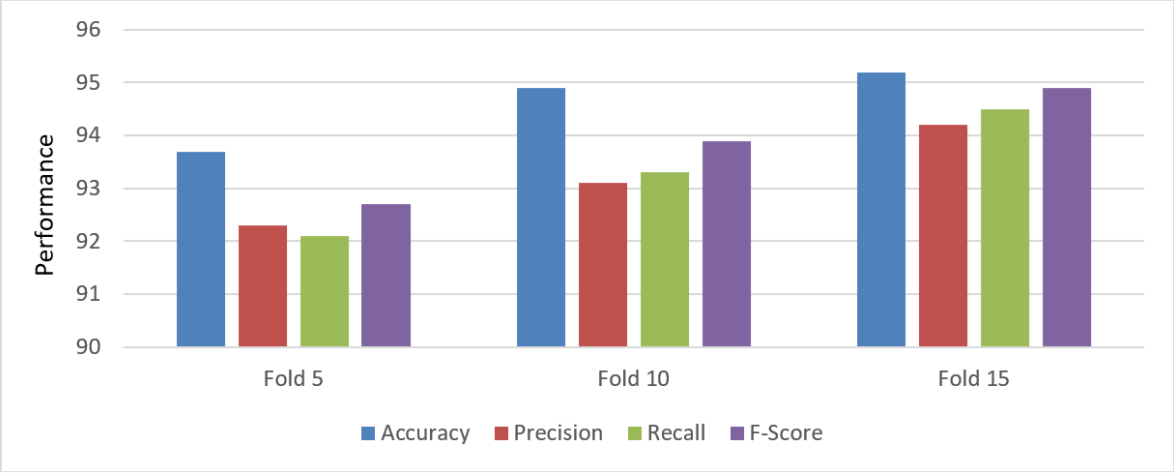

Figure 3. Validation of the model utilizing RNN-LSTM (sigmoid) classification models with 5, 10, and 15 folds of cross validation.

According to the results of the experiments, the 15-fold cross-validation method achieved the highest average classification accuracy of 95.10%.

4.2. Experiment with RNN-LSTM algorithm (Tanh)

In this experiment including the RNN-LSTM (Tan h) model, the accuracy, precision, recall, and f-score with various forms of cross validation are achieved. Consider the **Figure 4**, which presents the results of

validating the model with 5, 10, and 15 folds of cross validation utilising the RNN-LSTM (tan h) classification model.

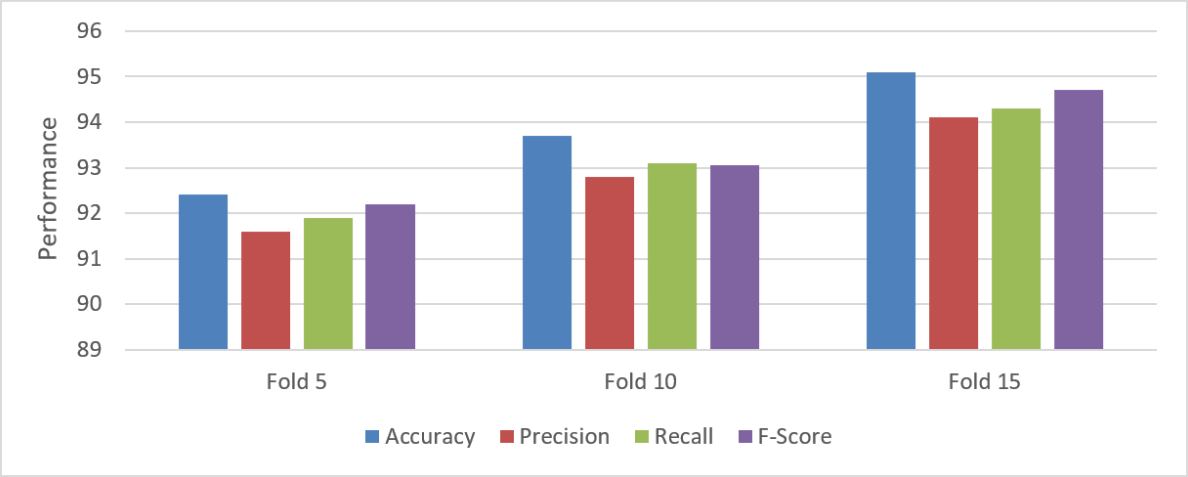

Figure 4. Validation of the model utilizing RNN-LSTM (Tanh) classification models with 5, 10, and 15 folds of cross validation.

According to the results of the experiments, the 15-fold cross-validation method achieved the highest average classification accuracy of 94.9%.

4.3. Experiment with RNN-LSTM algorithm (ReLU)

In this experiment including the RNN-LSTM (ReLU) model, the accuracy, precision, recall, and f-score with various forms of cross validation are achieved. Consider the **Figure 5,** which presents the results of validating the model with 5, 10, and 15 folds of cross validation utilising the RNN-LSTM (ReLU) classification model.

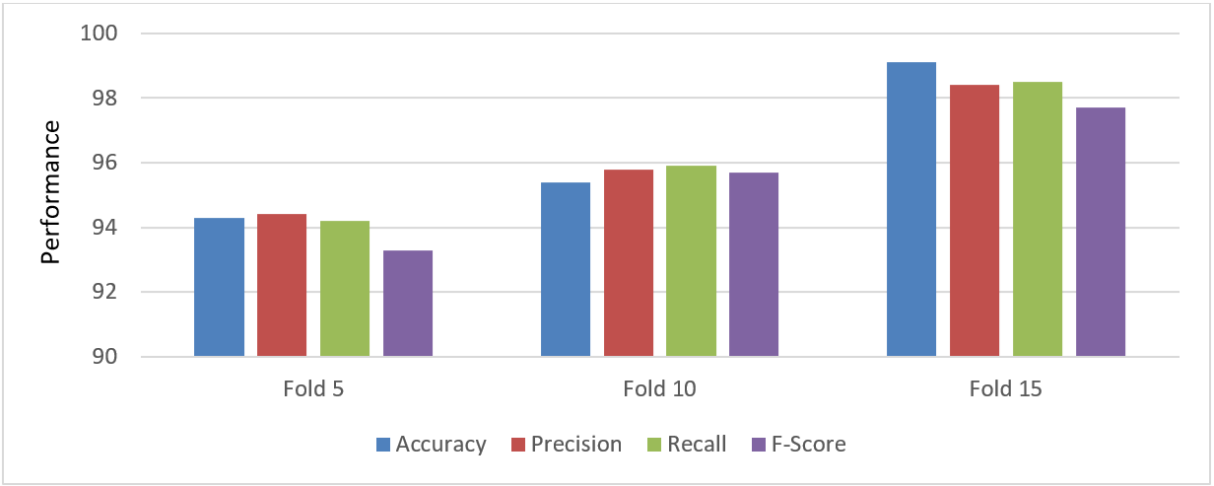

Figure 5. Validation of the model utilizing RNN-LSTM (Tanh) classification models with 5, 10, and 15 folds of cross validation.

According to the results of the experiments, the 15-fold cross-validation method achieved the highest average classification accuracy of 99.1%.

A total of at least three hidden layers were utilised in the investigation into the spread of fake news. Based on the findings of the empirical research, the conclusion can be drawn that the RNN-LSTM with ReLu function provides more accurate detection than both the RNN-LSTM (Tanh) function and the RNN-LSTM (sigmoid) function.

5. Conclusion and future scope

In the modern era, billions of people are always engaged in the usage of social media networks, and the total number of users on such platforms continues to rise on a daily basis. The ease with which people all around the world are able to consume misleading information on a regular basis is certainly cause for concern, and there is no denying this fact. As a consequence of this, in order to prevent this fake news from spreading further among consumer assets, a unique way for recognizing and reporting fake news is provided which utilizes RNN-LSTM model and makes use of the LIAR dataset. The process of extracting hybrid features from textual material, such as N-gram, TF-IDF, bigram, and dependency-based natural language processing characteristics is developed as a method. When compared to other traditional approaches to machine learning classifier, the RNN-LSTM method that was proposed obtains a greater level of accuracy than those other approaches. It achieves an accuracy rate of 99.10% both during training and testing, which is better to the accuracy achieved by common machine learning approaches such as SVM, ANN, RF, NB, and DT. In other words, it outperforms the accuracy achieved by these conventional methods. In the proposed approach for detecting false news using RNN-LSTM, the three experiments are conducted to acquire, accuracy, precision, recall, and f-score with varying forms of cross validation (5-Fold, 10-Fold and 15-Fold). Based on the findings of the empirical research, the conclusion can be drawn that the RNN-LSTM with ReLu function provides more accurate detection than both the RNN-LSTM (Tan h) function and the RNN-LSTM (sigmoid) function with -15-Fold cross validation. The future step in the development of this system will be to create a deep learning classifier that uses several consensus mechanisms to analyze the massive amounts of streaming data that take place in real time. The research that will be done in the future on this system will concentrate on this aspect.

Author contributions

Conceptualization, AKS and SS; methodology, AKS; software, AKS; validation, AKS, SS and BSK; formal analysis, AKS; investigation, AKS; resources, AKS; data curation, AKS; writing—original draft preparation, AKS; writing—review and editing, AKS; visualization, AKS; supervision, AKS; project administration, AKS. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Wynne HE, Swe KT. Fake news detection in social media using two-layers ensemble model. In: Proceedings of the 2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC); 5–8 July 2022; Phuket, Thailand. pp. 411–414.
2. Krishna NLSR, Adimoolam M. Fake news detection system using decision tree algorithm and compare textual property with support vector machine algorithm. In: Proceedings of the 2022 International Conference on Business Analytics for Technology and Security (ICBATS); 16–17 February 2022; Dubai, United Arab Emirates. pp. 1–6.
3. Mladenova T, Valova I. Research on the ability to detect fake news in users of social networks. In: Proceedings of the 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA); 9–11 June 2022; Ankara, Turkey. pp. 1–4.
4. Renuka M, Anithaashri TP. Enhancing the detection of fake news in social media using support vector machine algorithms comparing over apriori algorithms. In: Proceedings of the 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS); 10–12 October 2022; Tashkent, Uzbekistan. pp. 51–54.
5. Gupta V, Mathur RS, Bansal T, Goyal A. Fake news detection using machine learning. In: Proceedings of the 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON); 26–27 May 2022; Faridabad, India. pp. 84–89.
6. Tian Z, Baskiyar S. Fake news detection using machine learning with feature selection. In: Proceedings of the 2021 6th International Conference on Computing, Communication and Security (ICCCS); 4–6 October 2021; Las Vegas, NV, USA. pp. 1–6.

7. Agarwal P, Reddivari S, Reddivari K. Fake news detection: An investigation based on machine learning. In: Proceedings of the 2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI); 9–11 August 2022; San Diego, CA, USA. pp. 61–62.
8. Chaturvedi S, Saxena A, Gupta A, et al. Detecting fake news using machine learning algorithms. In: Proceedings of the 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N); 17–18 December 2021; Greater Noida, India. pp. 212–216.
9. Choudhary M, Chouhan SS, Pilli ES, Vipparthi SK. BerConvoNet: A deep learning framework for fake news classification. Applied Soft Computing 2021; 110: 107614. doi: 10.1016/j.asoc.2021.107614
10. Zervopoulos A, Alvanou AG, Bezas K, et al. Deep learning for fake news detection on Twitter regarding the 2019 Hong Kong protests. Neural Computing and Applications 2022; 34(2): 1–14. doi: 10.1007/s00521-021-06230-0
11. Kumar S, Arora B. A review of fake news detection using machine learning techniques. In: Proceedings of the 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC); 4–6 August 2021; Coimbatore, India. pp. 1–8.
12. Saleh H, Alharbi A, Alsamhi SH. OPCNN-FAKE: Optimized convolutional neural network for fake news detection. IEEE Access 2021; 9: 129471–129489. doi: 10.1109/ACCESS.2021.3112806
13. Goonathilake MDPP, Kumara PPNV. CNN, RNN-LSTM based hybrid approach to detect state-of-the-art stance-based fake news on social media. In: Proceedings of the 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer); 4–7 November 2020; Colombo, Sri Lanka. pp. 23–28.
14. Cao Z, Ding X, Tao Z. News detection for recurrent neural network approach. In: Proceedings of the 2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI); 23–25 September 2022; Nanjing, China. pp. 794–797.
15. Ghinadya, Suyanto S. Synonyms-based augmentation to improve fake news detection using bidirectional LSTM. In: Proceedings of the 2020 8th International Conference on Information and Communication Technology (ICoICT); 24–26 June 2020; Yogyakarta, Indonesia. pp. 1–5.
16. Krishna NLSR, Adimoolam M. Fake news detection system using logistic regression and compare textual property with support vector machine algorithm. In: Proceedings of the 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS); 7–9 April 2022; Erode, India. pp. 48–53.
17. Mahmud FB, Rayhan MMS, Shuvo MH, et al. A comparative analysis of graph neural networks and commonly used machine learning algorithms on fake news detection. In: Proceedings of the 2022 7th International Conference on Data Science and Machine Learning Applications (CDMA); 1–3 March 2022; Riyadh, Saudi Arabia. pp. 97–102.
18. Billones PJR, Macasaet DD, Arenas SU. Bilingual fake news detection algorithm using naïve bayes and support vector machine models. In: Proceedings of the 2022 IET International Conference on Engineering Technologies and Applications (IET-ICETA); 14–16 October 2022; Changhua, Taiwan. pp. 1–2.
19. Vinothkumar S, Varadhaganapathy S, Ramalingam M, et al. Fake News detection using SVM algorithm in machine learning. In: Proceedings of the 2022 International Conference on Computer Communication and Informatics (ICCCI); 25–27 January 2022; Coimbatore, India. pp. 1–7.
20. Divadkar S, Sahu A, Puri S. A review of ambiguous news detection approaches with deep learning, machine learning, and ensemble paradigms. In: Proceedings of the 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT); 7–9 October 2022; Bangalore, India. pp. 1–9.
21. Simone S, Oliver K, Clara D, Andreas S. Detecting fake news and performing quality ranking of German newspapers using machine learning. In: Proceedings of the 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME); 16–18 November 2022; Maldives, Maldives. pp. 1–5.
22. Gupta A, Mishra V, Briskilal J. Fake news detection using ensemble model. In: Proceedings of the 2022 Second International Conference on Interdisciplinary Cyber Physical Systems (ICPS); 9–10 May 2022; Chennai, India. pp. 34–39.
23. Umer M, Imtiaz Z, Ullah S, et al. Fake news stance detection using deep learning architecture (CNN-LSTM). IEEE Access 2020; 8: 156695–156706. doi: 10.1109/ACCESS.2020.3019735
24. 24. Han W, Mehta V. Fake news detection in social networks using machine learning and deep learning: Performance evaluation. In: Proceedings of the 2019 IEEE International Conference on Industrial Internet (ICII); 11–12 November 2019; Orlando, FL, USA. pp. 375–380.
25. Sahoo SR, Gupta BB. Multiple features-based approach for automatic fake news detection on social networks using deep learning. Applied Soft Computing 2021; 100: 106983. doi: 10.1016/j.asoc.2020.106983