International Journal of Information Management Data Insights 1 (2021) 100007

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

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

Fake news detection: A hybrid CNN-RNN based deep learning approach

Jamal Abdul Nasir[a,b], Osama Subhani Khan[b], Iraklis Varlamis[c,*]

[a] Data Science Institute, National University of Ireland Galway, Ireland
[b] Department of Computer Science, International Islamic University Islamabad, Pakistan
[c] Department of Informatics & Telematics, Harokopio University of Athens, Greece

ARTICLE INFO

Keywords:
Deep learning
Fake news detection
Misinformation
Disinformation
Rumours
CNN-RNN

ABSTRACT

The explosion of social media allowed individuals to spread information without cost, with little investigation and fewer filters than before. This amplified the old problem of fake news, which became a major concern nowadays due to the negative impact it brings to the communities. In order to tackle the rise and spreading of fake news, automatic detection techniques have been researched building on artificial intelligence and machine learning. The recent achievements of deep learning techniques in complex natural language processing tasks, make them a promising solution for fake news detection too. This work proposes a novel hybrid deep learning model that combines convolutional and recurrent neural networks for fake news classification. The model was successfully validated on two fake news datasets (ISO and FA-KES), achieving detection results that are significantly better than other non-hybrid baseline methods. Further experiments on the generalization of the proposed model across different datasets, had promising results.

1. Introduction

With the rapid advancement in the field of artificial intelligence, a large number of experiments continue to be conducted in order to solve problems which were never considered in the context of computer science. One such problem is that of fake news detection (Kumar and Shah, 2018; Pierri and Ceri, 2019). Since the access to news media has become very convenient, as soon as some noteworthy event occurs anywhere around the globe, different news sources tend to make their news eye catching in order to deliver news to as many people as possible on the worldwide web. This in turn results in the quick dissemination of news to millions and billions of people through a large number of news sources such as news channels, articles, websites, and social networking sites.

Apart from several reputable news organizations and agencies which have been operating on an international level for decades, and deliver news to the general public, there are a large number of smaller news sources which deliver news that are not trustworthy. In addition to this, in popular social networking and social media platforms anyone from anywhere around the globe can publish and disseminate any kind of statement, or set of statements, to spread fake news through the use of different networking sites in order to achieve different goals, which may be fraudulent or illegal. A major caveat is that some of the sources that are considered to be authentic, and are popular sources for informational services, such as Wikipedia, are also prone to false information or fake news (Kumar et al., 2016). In addition to this, because some official news aggregators may deliberately spread false or fake news in order to gain popularity, achieve some political objective, or earn money, the problem is further intensified. Another factor contributing to the spread of fake news may be organized astroturfing campaigns which attempt to mock or spoil a specific product or company, a society or a group of people, e.g. for political, social, or financial reasons (Ratkiewicz et al., 2011).

Fake news is considered to be one of the greatest threats to commerce, journalism and democracy all over the world, with huge collateral damages. A US $130 billion loss in the stock market was the direct result of a fake new report that US president Barak Obama got injured in an explosion.[1] Other cases of fake news campaigns that demonstrate the enormous impact that fake news can have include the sudden shortage of salt in Chinese supermarkets after a fake report that iodized salt would help counteract the effects of radiation after the Fukushima nuclear leak in Japan,[2] and an escalation of tensions between India and Pakistan that

[1] Rapoza K., Can 'fake news' impact the stock market? In Forbes. 26 February, 2017. Available online at: https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market.

[2] Wang Jingqiong and Li Xinzhu, Radiation fears prompt panic buying of salt. In China Daily. 18 March, 2011. Available online at: http://www.chinadaily.com.cn/2011-03/18/content_12189516.htm.

* Corresponding author.
E-mail address: varlamis@hua.gr (I. Varlamis).

https://doi.org/10.1016/j.jjimei.2020.100007
Received 23 November 2020; Received in revised form 28 December 2020; Accepted 29 December 2020

began with fake reporting of the Balakot strike and resulted in the deaths of military personnel and the loss of expensive military equipment.[3]

The term 'fake news' is often described in related literature as 'misinformation', 'disinformation', 'hoax', and 'rumor', which are actually different variations of false information. There are a variety of research projects, tools and applications for fact checking, and fake news detection (Zubiaga et al., 2018), which mostly examine the problem as a veracity classification.

Misinformation is used to refer to the spreading of false information disregarding the true intent. False information can be the result of false labeling (e.g. of persons in a photo), poor fact-checking, etc., and can easily spread among users that do not care much about the veracity of what they are reading or sharing. From an etymological point of view, the term combines the prefix mis', which means "wrong" or "mistaken", with information. *Disinformation* implies an intent to mislead the target of information. It refers to false information that is disseminated in a tactical way in order to bias and manipulate facts. It is usually coined with the term "propaganda". The prefix dis- is used to indicate a reversal or negative instance of information. *Rumours* and *hoaxes* are interchangeably used to refer to false information that is deliberately constructed to seem true. The facts that they report are either inaccurate or false, although they are presented as genuine.

The dissemination of fake news through different medium especially online platforms has not been stopped completely or scaled down to a degree in order to reduce the adverse effects fake news can lead to. The reason is that there is no system that exists that can control fake news with little or no human involvement. Experiments indicate that machine and learning algorithms may have the ability to detect fake news, given that they have an initial set of cases to be trained on.

Deep learning techniques have great prospect in fake news detection task. There are very few studies suggest the importance of neural networks in this area. The model proposed is the hybrid neural network model which is a combination of convolutional neural networks and recurrent neural networks. As this model is required to classify between fake news and legitimate news, so this problem is cast as a binary classification problem.

Section 2 that follows perform a survey of fake news datasets, tasks, algorithms and major research works. Section 3 presents the proposed approach and its main components. Section 4 presents the implementation details of this work and explains the experimental evaluation setup. Section 5 presents the results of this evaluation and Section 6 illustrates the performance of the proposed method and its alternatives. Finally, Section 7 discusses the contribution of this work to the theory and practice of deep learning for fake news detection and Section 8 concludes the paper and summarizes the next steps.

2. Literature review

Research in the area of news verification and debunking of false information on the web, involves different activities in order to achieve best results (Rashkin et al., 2017). Despite the rapid increase in its popularity, the subject is still in its infancy among researchers. While there has been an increase in the number of studies that focus on the analysis and study of fake news and/or rumor characteristics in order to properly identify and debunk false information, there is still enough room left for research in this direction since there is not a unified solution yet (Vosoughi et al., 2018).

Another important task, in the case of misinformation and disinformation in social networks, is to study the way false or fake information spreads over the network (Wang et al., 2019) and early detect suspicious dissemination patterns (e.g. spamming Sahoo et al., 2020; Wu et al., 2019), astroturfing (Keller et al., 2020) etc. Concerning misinformation, a recent study on its diffusion trends in various social media across different periods (Allcott et al., 2019) shows that misinformation increases before major political events. According to Aswani et al. (2019), the use of emotion and polarity of the associated text can help in determining the authenticity of shared information.

2.1. Related datasets

One of the earliest works on the automatic detection of fake news was by Vlachos and Riedel (2014). Authors defined the task of fact-checking, collected a dataset from two popular fact-checking websites and considered k-Nearest Neighbors classifiers for handling fact-checking as a classification task. Wang (2017) released the LIAR dataset, which comprises 12.8K manually labelled short statements from PolitiFact. Several classifiers were compared on a six truthfulness levels classification task and achieved better results by incorporating additional textual and contextual features such as subject, speaker, history, etc.

Ma et al. (2016) collected 5 million posts from Twitter and Sina Weibo micro blogs that comprised 778 reported events from which 64% are rumors. No feature engineering techniques were applied to classify events using several baseline algorithms (i.e. Decision Tree, Random Forest, SVMs) as well as RNNs and LSTMs. Another work on the same dataset proposed a hybrid deep learning model, CSI which is composed of three modules namely capture, score and integrate (Ruchansky et al., 2017). The first captures the temporal pattern of user activity based on the response of users to text, using an LSTM RNN. The second captures source characteristics by implementing a fully-connected neural network layer. The last module combines the other two in the last layer for detection. They achieved 89% and 95% accuracies on Twitter and Weibo datasets, respectively.

CREDBANK (Mitra and Gilbert, 2015) is a large-scale data set, comprising 37 million event related Tweets, grouped into over 1,000 sets (events). Events have been manually annotated for their accuracy level (5-point Likert scale). The EU project PHEME (Derczynski et al., 2015) has released several rumor related datasets that mostly challenge the ability to evaluate the veracity of news. The *FakeNewsNet* dataset was released by Shu et al. (2018) and is updated on a daily basis with new content. It combines claims from PolitiFact, GossipCop with social context and spaciotemporal information for each content. Authors evaluated the performance of SVM, LR, Nave Bayes (NB) and CNN in the dataset.

The FA-KES dataset comprises news events around the Syrian war (Salem et al., 2019). The dataset consists of 804 news articles of which 376 are fake. Annotations were semi-supervised using a fact-checking labelling approach that employed the Syrian Violation Documentation Center (VDC) database. Although the dataset can be used to train machine learning models for fake news detection for other related domains than war-related news it is rather small.

A publicly available dataset for stance classification of rumored claims, is 'Emergent' (Ferreira and Vlachos, 2016). The data-set contains 300 rumored claims and 2,595 associated news articles, collected and labelled by journalists with an estimation of their veracity (true, false or unverified). Authors addressed the task of determining the article headline stance with respect to the claim. The claims are collected from a variety of sources such as rumor sites (e.g. Snopes), and Twitter accounts (e.g. @Hoaxalizer).

A large-scale dataset, namely FEVER (Thorne et al., 2018), generated claims from Wikipedia articles and classified them to three classes (Supported, Refuted or NotEnoughInfo). The proposed system for Fact Extraction and VERification, is a pipeline approach that involves Document Retrieval, Sentence Selection and Recognizing Textual Entailment tasks.

In an effort to analyse tree-structured conversation formed of tweets replying to an original rumourous tweet, the SemEval-2017, Task 8 (RumourEval) (Derczynski et al., 2017) and its continuation SemEval-2019,

[3] BBC News, India and Pakistan: How the war was fought in TV studios. In BBC news. 10 March, 2019. Available online at:https://www.bbc.com/news/world-asia-47481757.

Task 7 (RumourEval 2019) released two Twitter based datasets for rumour evaluation. Although the datasets are publicly available, they contain only tweets and are limited to social media rumors.

2.2. Methods and models

According to the survey of Li et al. (2016), the "discovery of truth" largely focused: i) on the detection of subject-predicate-object triples, which represent the structured facts that are assessed for their credibility, and ii) on the classification of general text input using neural networks trained on labeled texts from sites like PolitiFact.com. When no external evidence or user feedback is available there is limited context for credibility analysis. When external evidence exists in the form of articles that confirm or refute a claim, the assessment of sources' trustworthiness, and claims' credibility is possible using supervised classifiers. The later approach requires substantial feature modeling and rich lexicons to detect bias and subjectivity in the language style.

DeClarE is an end-to-end neural network model proposed by Popat et al. (2018) for debunking fake news and false claims. It employs evidences and counter-evidences extracted from the web to support or refute a claim. Without feature engineering and manual interventions authors achieved an overall 80% classification accuracy on four different datasets, by training a bi-directional LSTM model with attention and source embeddings.

For automatically identifying fake news on twitter, Buntain and Golbeck (2017) employed structural, user, content and temporal features, a feature selection process and a Random Forest (RF) classifier. They evaluated their approach on CREDBANK and Pheme datasets. They also showed that models trained against crowdsourced workers outperform models based on journalists' assessment and models trained on a pooled dataset of both crowdsourced workers and journalists.

The TI-CNN (Text and Image information based Convolutional Neural Network) model has been proposed in Yang et al. (2018). TI-CNN is trained with both the text and image information simultaneously. The convolutional neural network makes the model to see the entire input at once, and it can be trained much faster than LSTM and many other RNN models. The dataset that focuses on the American presidential election consists of 20,000 news with almost 12,000 fake news.

Karimi and Tang (2019) provided a new framework for fake news detection. The framework learns the Hierarchical Discourse-level Structure of Fake news (HDSF), which is a tree-based structure that represents each sentence separately. Authors evaluated the framework on a merged dataset. They represented documents as N-gram and Linguistic Inquiry and Word Count (LIWC) vectors as well as vectors of Rhetorical Structure Theory (RST) relations, and employed SVM, LSTM and a hybrid Bi-directional Gated Recurrent Neural Network (BiGRNN) and CNN model for classification. HDSF outperformed all other methods with an 82.19% accuracy.

Ahmed et al. (2018) performed experiments in two different datasets comprising fake news and fake reviews. They utilized different variations of Term Frequency (TF) and TFIDF for feature extraction and SVM, Lagrangian-SVM (LSVM), KNN, DT, Stochastic Gradient Decent (SGD) and LR classifiers.

The theory-driven model of Zhou et al. (2020) studied the content-based and propagation-based characteristics of fake news. In order to detect fake news before its propagation, they provided a detailed analysis of the properties and characteristics of content-based and propagation-based methods. News content has been analysed at lexicon-, syntax-, semantic- and discourse-level. Also, features related to deception/disinformation and clickbaits, and the impact of news distribution were studied. They evaluated SVM, RF, Gradient Boosting (XGB), LR and NB on the FakeNewsNet dataset.

The benchmark study of Khan et al. (2019) evaluated SVM, LR, DT, Adaboost, NB, KNN, CNN, LSTM, Bi-LSTM, Conv-LSTM, Hierarchical Attention Network (HAN), Convolutional-HAN and character-level C-LSTM classifiers on three different balanced datasets. They showed that although neural networks perform better on larger datasets, NB can perform as good as neural networks on smaller datasets. Finally, Elhadad et al. (2019) experimented using hybrid sets of features extracted from online news content and textual metadata on three publicly available datasets (ISOT, FA-KES and LIAR). Using DT, KNN, LR, SVM, Bernoulli NB, multinomial NB, LSVM, perceptron and neural networks they achieved a 100% accuracy on ISOT dataset, 62% accuracy on LIAR dataset and a benchmark accuracy of 58% on FA-KES dataset.

Hybrid methods are frequently employed for fake news detection. Ajao et al. (2018) have tested an LSTM-CNN variation, which included a 1D CNN immediately after the word embedding layer of the LSTM model, and achieved an accuracy of 0.80 in fake tweets prediction. Hybrid LSTM models that train the attention mechanism using local semantic (topics) or user (profile) attentions have also been proposed (Long, 2017). The hybrid deep learning model of Ruchansky et al. (2017) combines temporal representations of articles using an RNN, with a fully connected layer that aggregates features from the users that reproduce the news. The hybrid model of Hamdi et al. (2020) combines graph embeddings from the user follower graph on Twitter with user features in order to evaluate the credibility of sources and thus of the news they publish or share. The hybrid CNN model of Wang (2017) uses speaker profiles for training.

Among the various hybrid methods that exist in the literature, those that model the social graph that spreads the news, or the user and news source features (profile), cannot be applied when only the text of the news is available. From the hybrid methods that examine only the textual content of news, the combination of LSTM and CNN has shown promising results. However, so far, LSTMs have been used for providing word embeddings and CNN for doing the final classification.

2.3. Related tasks

The importance of identifying and debunking fake news triggered the importance of stance classification task. ***Stance classification*** involves determining the stance of certain text against a claim. Consequently, plays an important role in fake news detection, since texts or personal statements that are against a claim may lead to a false classification of the claim as fake news (Zubiaga et al., 2018).

A stance detection challenge was conducted by SemEval[4] under the name "Detecting Stance in Tweets". Another dataset for stance classification of rumored claims, called 'Emergent' was presented in Ferreira and Vlachos (2016). The data-set contains 300 rumored claims and 2,595 associated news articles, collected and labelled by journalists with an estimation of their veracity (true, false or unverified). Each associated article is summarized into a headline and labelled to indicate whether its stance is for, against, or observing the claim. Authors addressed the task of determining the article headline stance with respect to the claim. The claims in the Emergent dataset are collected by journalists from a variety of sources such as rumor sites, e.g. Snopes, and Twitter accounts such as Hoaxalizer. Their subjects include topics such as world and national U.S. news and technology stories. The models developed own features and used a Logistic Regression (LR) classifier to measure the agreement between headline and the claim. They reported an accuracy of 73% of their classifier with Emergent. However, the dataset collected is rather small to learn all the nuances of the task.

The classification model presented by Pamungkas et al. (2019) outperformed all other models in the Semeval-2017 Task 8, by using conversation-based and affective-based features covering different facets of affect.

The Maester framework (Shang et al., 2018) was proposed for the evaluation of news content. Its goal was to display results against a search query with articles classified as agreeing articles, disagreeing ar-

[4] http://alt.qcri.org/semeval2016/task6/.

ticles and discussing articles. This can be considered as an application of stance detection.

Research in ***Rumours*** is limited to social media. Rumor research differs from fake news research, but it has similar characteristics. An unverified information circulating in social media constitutes a rumor which can or cannot be substantiated or debunked later with time; it is considered as an unconfirmed information rather than a hoax or false information that is deemed false (Zubiaga et al., 2018; 2017).

The research related to rumors in social media is still in its early stages. Although, the literature is scarce in this domain, there has been an increase in the interest of rumors in social media (Ma et al., 2015; Zubiaga and Ji, 2014). A complete outline of architecture of a rumor classification system is proposed by Zubiaga and Ji (2014), and consists of four stages: (1) rumor detection; (2) rumor tracking; (3) rumor stance classification and (4) rumor veracity classification. Much of the work done is focused more on stance classification of rumors (Lukasik et al., 2015; Zeng et al., 2016; Zubiaga et al., 2016). Rumor veracity classification is also much focused by researchers in this domain (Jin et al., 2016; Zubiaga and Ji, 2014).

A conclusion to the analysis of the related literature, is that fake news has played a significant role in many real time disasters and resulted in severe consequences for journalism, economy and political instability. In order to tackle fake news, manual interventions are of no use due to the fast-paced information sharing on the internet. Machine learning techniques have been experimented on a range of datasets and deep learning techniques are still to be fully evaluated on the fake news detection and related tasks.

3. Proposed approach

The research on fake news detection requires a lot of experimentation using machine learning techniques on a wide range of datasets. Novel techniques must get deep understanding of the nature of fake news and the ways they spread around the world. The current work contributes in this direction by proposing a model based on novel techniques that prove the importance of deep learning models for the fake news detection task. More precisely, it introduces a combination of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), which boosts the performance of the proposed fake news detection model.

Neural Networks are the most widely used type of bio-inspired computational methods nowadays. Their performance in several classification tasks is the state of the art in each field. More recent developments in bio-inspired computing have introduced new algorithms (Kar, 2016) which have been employed successfully in detection tasks that are highly related to fake news. Examples include the detection of spammers (Aswani et al., 2018), sentiment analysis (Yadav and Vishwakarma, 2020), etc.

Bio-inspired computational techniques demonstrate many applications in optimisation, search and scheduling problems, whereas deep neural networks are usually applied in classification and prediction tasks (Kar, 2016). In this work, fake news detection is modelled as a classification task, which makes deep neural networks more appropriate. This does not limit the potential of other bio-inspired techniques to work on a different formulation of the problem that also considers information diffusion paths and incorporates information about nodes that generate and reproduce news (Meel and Vishwakarma, 2019). When it comes to detection of fake images and fact-checking based on image analysis, deep learning techniques, and CNNs specifically, have been proven very successful, since they allow face recognition and classification (Bouchra et al., 2019), image segmentation, object detection and characterisation (Dhillon and Verma, 2020; Rajagopal et al., 2020) etc.

In order to cover all the research aspects and justify the legitimacy, validity and reliability of this research the following actions are taken:

- First, an extensive literature review is performed, using Google Scholar search and a search on related GitHub repositories, in order to identify the relevant publications and experiments on different datasets and propose a new model that fills existing gaps.
- Then the proposed solution is introduced, which aims to fill the identified research gap: “the lack of thorough investigations on relevantly new datasets and the lack of combinations of deep learning models or neural networks for fake news detection”.
- The novel hybrid deep learning model is evaluated on a publicly available dataset and compared with state of the art approaches on the same dataset in order to justify its validity. The results produced by Elhadad et al. (2019) on the FA-KES dataset is the comparison base. Additional baselines are produced with the help of a total of seven supervised machine learning classification techniques, and allow further validation of the proposed model.
- Finally, further experimentation is performed on the ISOT dataset (Ahmed et al., 2018) in order to state the possible future work in this research direction and reach to a conclusion.

3.1. Neural Networks and decisions

Neural Networks take their input in the form of numerical vectors or matrices and perform calculations, such as additions and matrix multiplication in order to tune their parameters. There are different parameters that affect the performance of a neural network, such as the initialization of weights and biases, the activation functions used at each layer, the optimizer and a loss function.

The purpose of the activation functions on a layer is to use weights, biases and a given input and provide output for the next layer. The final layer is the output layer that gives the results. A loss function determines the error between the computed and the desired result for each input. In order to reduce the error between the computed and the desired result, a neural network uses the optimizer. The weights are randomly initialized and are then trained with the help of an optimization technique. There is a large variety of optimizers, activation functions and loss functions to choose. In the above, one has to add the different types of layer and connection between layers, which results in a multitude of neural and deep neural network architectures (Liu et al., 2017).

In the case of fake news detection, the task can be considered a binary classification task. The two outputs denote that a piece of information is either fake or real.

3.2. Word embeddings

When dealing with text classification and neural networks, the input text must take a vector or matrix numeric format so that it can be fed to the network. Words in the text can be represented as vectors, which are referred to as word vectors with each word having a unique word vector. These word vectors are referred to as word embeddings.

Word embeddings are trained from a large corpus, which is usually language specific, or domain specific in order to allow the vocabulary of embeddings capture the statistical relations of all the words in the corpus. Instead of training word embeddings, it is more feasible to use publicly available pre-trained word embeddings. The most popular pre-trained word embeddings available are Word2Vec provided by Google, and GloVe. From the words in the text, those that are found in the vocabulary are kept, whereas those that are not in the vocabulary are ruled out.

3.3. Convolutional Neural Network

A Convolutional Neural Network (CNN) involves multiplication of matrices that provide outputs to incorporate for further training process. This method is known as convolution. That is why this type of neural network is called a convolutional neural network. In the case of NLP, words in a sentence or a news article are represented as word vectors.

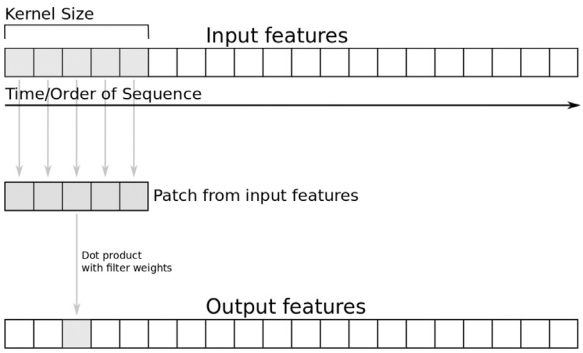

Fig. 1. An 1-D convolutional operation.

These word vectors are then used for training a CNN. The training is carried out by specifying a kernel size and a number of filters. A CNN can be multi-dimensional.

In the case of text classification or NLP, a one-dimensional CNN (Conv1D) is used generally. Conv1D deals with one dimensional arrays representing word vectors. In a CNN, a filter of fixed size window iterates through the training data, which at each step multiplies the input with the filter weights and gives an output that is stored in an output array. This output array is a feature map or output filter of the data. In this way a feature is detected from the input training data. This process can be visualized in Fig. 1.

The size of the filter is specified as kernel size and the number of filters specifies the number of feature maps to be used. In this way CNN can be utilized to learn local features that are directly derived from the training data.

3.4. Recurrent Neural Networks

A Recurrent Neural Network (RNN) involves sequential processing of the data for learning. This sequential process is justified by its ability to retain a memory of what came before the current sequence being processed. It is called recurrent because the output at each time step is utilized in the next time step as input. This is done by remembering the output of the previous time step. This in turn allows us to learn long-term dependencies in the training data.

In the case of NLP, many news articles can be considered for learning relative to each other instead of separately learning each news article. RNN is composed of layers with memory cells. There are different types of memory cells to utilize in RNN. One such type is the Long Short-Term Memory (LSTM) unit or cell (Graves, 2012). LSTM consists of a cell state and a carry in addition to the current word vector in process as the sequence is processed at each time state. The carry is responsible to ensure that there is no information loss during the sequential process.

The anatomy of an LSTM cell is shown in Fig. 2. An LSTM cell is composed of weights and three different gates for learning. At each time step, there is an input gate for the current input, an output gate to predict values and a forget gate that is used to discard information that is irrelevant.

3.5. Hybrid CNN-RNN model

The proposed model makes use of the ability of the CNN to extract local features and of the LSTM to learn long-term dependencies. First, a CNN layer of Conv1D is used for processing the input vectors and extracting the local features that reside at the text-level. The output of the CNN layer (i.e. the feature maps) are the input for the RNN layer of LSTM units/cells that follows. The RNN layer uses the local features extracted by the CNN and learns the long-term dependencies of the local features of news articles that classify them as fake or real. The proposed model is depicted in Fig. 3.

The combination of CNN-RNN has been proven successful in several classification and regression tasks, since they have the ability to capture both local and sequential characteristics of input data. For example, they have been used for emotion detection (Kollias and Zafeiriou, 2020) and sign language recognition from video streams (Masood et al., 2018), taking advantage of their ability to learn scene features using the CNN and sequential features using the RNN. In the case of NLP tasks, RNN can learn temporal and context feature from text, and capture long-term dependencies between text entities and important features, which are detected using the ability of CNN in handling spatial relations (Zhang et al., 2018; Zhou et al., 2015).

Despite their advantages, deep learning models have certain practical limitations, such as the difficulty in finding the optimal hyper-parameters for each problem and dataset, the requirement for big training datasets, and the lack of interpretability, which have a direct effect on their performance in new and unknown tasks and make them behave as black-box oracles (Drumond et al., 2019). Recent advances in bio-inspired methods, allow the optimisation of deep learning parameters and form the basis of next-generation optimization algorithms for machine learning. The proposed hybrid model can be significantly benefited from the optimisation of its hyper-parameters and is part of next work in this field to examine the various bio-inspired techniques (Kar, 2016) and find the most appropriate for the current task.

4. Implementation and evaluation

This work implements the aforementioned hybrid model in Python and evaluates its performance on two real-world fake news datasets. The following subsections describe the datasets in detail, explain the implementation decisions and the comparison baselines, as well as the achieved results. The baseline methods comprise state-of-the-art techniques for fake news classification and a few more newly produced baselines.

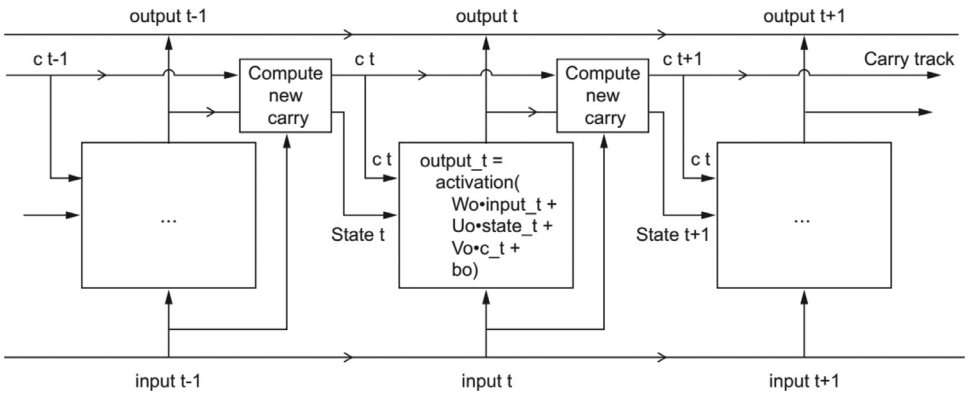

Fig. 2. The anatomy of an LSTM cell.

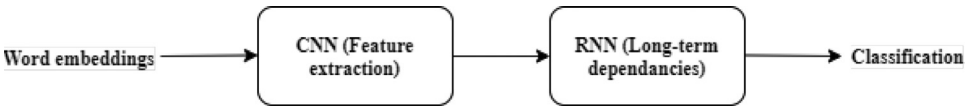

Fig. 3. The proposed hybrid model.

Table 1
Datasets that can be used for fake news and rumors detection as well as for fact checking.

Work	Source	URL	Detection
Vlachos and Riedel (2014)	PolitiFact Channel4	https://sites.google.com/site/andreasvlachos/resources	Fake news 2 classes
Mitra and Gilbert (2015)	CREDBANK	https://github.com/compsocial/CREDBANK-data	Credibility 5 levels
Derczynski et al. (2015)	PHEME project	https://www.pheme.eu/software-downloads/	Veracity
Mihaylova et al. (2019)	SemEval-19 T.8	https://competitions.codalab.org/competitions/20022	Fact Checking Q&A classes Rumours
Shu et al. (2018)	FakeNewsNet	https://github.com/KaiDMML/FakeNewsNet	Fake News
Salem et al. (2019)	FA-KES	https://zenodo.org/record/2607278#.X3oK8WgzaUk	Fake News
Ferreira and Vlachos (2016)	Emergent	https://github.com/willferreira/mscproject	Rumours
Thorne et al. (2018)	FEVER	http://fever.ai/	Fact Verification
Derczynski et al. (2017)	RumourEval	http://alt.qcri.org/semeval2017/task8/index.php?id=data-and-tools	Rumours
Gorrell et al. (2019)	RumourEval-2019	https://competitions.codalab.org/competitions/20022	Rumours Stance labels

Table 2
Techniques and datasets that have been used for fake news detection and fact checking.

Work	Source	URL	Detection	Model
Wang (2017)	PolitiFact **LIAR**	https://www.cs.ucsb.edu/~william/data/liar_dataset.zip	Fake news 6 levels	majority LR, SVM, bi-LSTM, CNN
Ma et al. (2016) Ruchansky et al. (2017)	Snopes (Twitter,Weibo)	http://alt.qcri.org/~wgao/data/rumdect.zip	Rumors 2 levels	RF, DT SVM, RNN LSTM
Popat et al. (2018)	Snopes PolitiFact NewsTrust SemEval-17 T.8	https://www.mpi-inf.mpg.de/dl-cred-analysis/	Credibility 2 or 5 levels	bi-LSTM, LSTM, CNN
Buntain & Gollbeck (2017)	CREDBANK PHEME	–	Credibility	RF
Yang et al. (2018)	Kaggle & News	https://drive.google.com/open?id=0B3e3qZpPtccsMFo5bk9Ib3VCc2c https://www.kaggle.com/mrisdal/fake-news	Fake news 2 classes	TI-CNN,LSTM,RNN
Karimi and Tang (2019)	BuzzFeed Politifact Kaggle	https://www.kaggle.com/mrisdal/fake-news/data https://www.kaggle.com/jruvika/fake-news-detection	Fake news 2 classes	N-grams, LIWC, RST, BiGRNN-CNN LSTM, HDSF
Ahmed et al. (2018)	Politifact TripAdvisor **ISOT**	https://www.uvic.ca/engineering/ece/isot/datasets/fake-news/index.php	Fake news & reviews	SVM, SDG, LR, kNN, DT
Zhou et al. (2020)	FakeNewsNet Politifact BuzzFeed	–	Fake news Clickbaits Disinform.	SVM, RF XGB, LR, NB
Elhadad et al. (2019)	ISOT FA-KES LIAR	–	Fake news	DT, kNN, LR, SVM, NB, LSVM,NN
Pamungkas et al. (2019)	Twitter	http://alt.qcri.org/semeval2016/task6/	Stance	-
Ferreira and Vlachos (2016)	Snopes Hoaxalizer (Twitter)	https://github.com/willferreira/mscproject	Stance	LR

4.1. Datasets

The evaluation experiments are performed on two publicly available datasets, which contain news articles in English, as explained in the following.

The **FA-KES**[5] dataset consists of 804 news articles on the Syrian war. For each article, the headline, date, location and news sources are also provided in addition to the full body of text. A class label with values '0' for fake news and '1' for real news is also available. A basic text classification algorithm can use only the article body. The 426 articles are true and the remaining 376 are fake, which corresponds to a well-balanced dataset (53% true versus 47% fake articles). Although deep learning models work better on larger datasets, this small dataset allows to experiment with various deep and machine learning techniques and quickly get insights on the most appropriate techniques and their configuration.

The **ISOT**[6] dataset consists of 45,000 news articles, almost equally distributed to the true and fake categories. The true articles were collected from the Reuters website, and the fake ones from various sources flagged as fake sources by Wikipedia[7] and from Politifact. The datasets comprises the full body of each article, the title, date and topic. The main article topics are politics and world news and the dates fall between 2016 and 2017. The distribution by types and topics is given in Table 3.

Table 3
The breakdown of the ISOT dataset.

News	Total number of articles	Type	Number of articles
Real	21417	World news	10,145
		Politics news	11,272
Fake	23481	Government news	1570
		Middle east	778
		US news	783
		Left news	4459
		Politics	6841
		News	9050

[5] https://zenodo.org/record/2607278#.X3oK8WgzaUk.
[6] https://www.uvic.ca/engineering/ece/isot/datasets/fake-news/index.php.
[7] https://en.wikipedia.org/wiki/List_of_fake_news_websites.

4.2. Implementation decisions

The hybrid deep learning model is implemented on Google colab[8]. Colab is a Jupyter[9] notebook cloud environment that provides GPUs and TPUs for heavy computation. The code for the experiments (pre-processing and ML classifiers) is written in Python.

More specifically, the Keras Python[10] package is used for implementing the hybrid CNN-RNN model. Pandas and Numpy packages are used for reading the datasets and processing arrays respectively. NLTK package is used for data pre-processing. Scikit-learn[11] package is used for processing the data, results evaluation and baseline classifiers implementation. Matplotlib package is used for plotting graphs.

4.2.1. Dataset splitting and pre-processing

The datasets are read as Pandas DataFrame objects and the class labels are encoded using scikit-learn's LabelEncoder. Next, each dataset is spilt into training and test subsets (80–20% split).

For the validation of the classification models, all datasets have been pre-processed, in order to convert the raw texts in the appropriate format for each models. A python script is written especially for this task. Texts are first cleared from IP and URL addresses, using the *re* python package for regular expressions. Then the test is split into sentences and terms English stopwords are removed and the remaining terms are stemmed using the NLTK package.

4.2.2. Mapping text to vectors using word embeddings

The label matrices for training and test sets are encoded and the text is vectorized by tokenization using the Tokenizer function of the Keras library. The task is repeated separately for training and test texts in the two datasets. The tokenizer is fitted on the pre-processed training corpus which is converted to sequences of integers. The length of sequences is set to 300 and a post-padding is applied in order to use it for model training. This is done because the length of each sequence varies. By fixing the length of each text sequence to 300, it is necessary to append zeros (zero values) in each sequence that is shorter than the fixed length.

In order for the CNN to perform local feature extraction, pre-trained word embeddings are used. An embedding matrix is prepared using the GloVe pre-trained word embeddings.[12] GloVe was trained with a dataset of six billion words or tokens using a vocabulary of 400 thousand words, and provides embeddings in a range of dimensions. Word embeddings of 100 dimensions are used in this research. The embeddings' matrix is prepared using only the words that occur in the tokenized training corpus.

4.2.3. Model implementation in Keras

The proposed hybrid deep learning model is implemented using the Sequential model of the Keras deep learning Python library. The Sequential model comprises several layers of neurons:

- The first layer of the neural network is the Keras embedding layer. This is the input layer through which the pre-trained word embeddings are utilized by providing the prepared embedding matrix and the model is trained by feeding in the training data.
- The next layer is the one-dimensional CNN layer (Conv1D) for extraction of local features by using 128 filters of size 5. The default Rectified Linear Unit (ReLU) activation function is used.
- After that, the large feature vectors generated by CNN are pooled by feeding them in to a MaxPooling1D layer with a window size of 2, in order to down-sample the feature vectors, reduce the amount of parameters, and consequently the computations without affecting the network's efficiency.

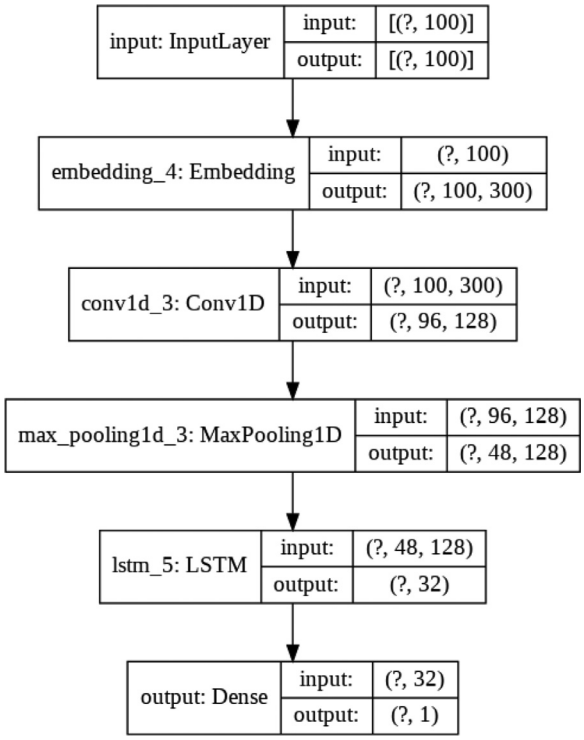

Fig. 4. FA-KES Model Summary.

- The pooled feature maps are fed into the RNN (LSTM) layer that follows. This input is used to train the LSTM, which outputs the long-term dependent features of the input feature maps, while retaining a memory. The dimension of the output is set to 32. The default linear activation function (i.e. $f(x) = x$) of Keras is used in this layer.
- Finally, the trained feature vectors are classified using a Dense layer that shrinks the output space dimension to 1, which corresponds to the classification label (i.e. fake or not fake). This layer applies the Sigmoid activation function.

The model is trained using the adaptive moment estimation (Adam) optimizer to define the learning rate in each iteration, the binary cross-entropy as the loss function, and the accuracy for the evaluation of results. The training is performed for 10 epochs using a batch size of 64.

Model summaries of FA-KES and ISOT datasets are shown in Figs. 4 and 5 respectively.

4.2.4. Model complexity

The time or computational complexity of Deep Neural Network (DNN) models is closely related to hardware execution, but is strongly related to the number of layers, the number of operations required to produce a result, and the number of elements to be processed. The computational complexity of Deep Neural Networks is an issue, especially for real-time tasks (Granger et al., 2017). Space complexity is closely related to model capacity, which for DNNs can be perceived as the number of parameters of the model in all layers.

The proposed hybrid model sequentially combines a CNN and an RNN model. As depicted in Figs. 4 and 5 it is composed by 6 layers, including the input (I), embedding (E), and output (O) layers. The total number of filters in the 1-D convolutional layer (128x5 sized kernel) is 128x5x100, where 100 is the length of the embeddings. The pooling layer is a fixed operation with no weighting factor.

The LSTM layer maintains a total of 4 sets of parameters, corresponding to the input gate, output gate, forgetting gate and candidate state. So the total number of parameters W ignoring the biases can be calculated as: $W = n_c \times n_c \times 4 + n_i \times n_c \times 4 + n_c \times n_o + n_c \times 3$, where n_c is the number of memory cells (n_c=48), n_i is the number of input units (n_i =128),

[8] https://colab.research.google.com/.
[9] https://jupyter.org/.
[10] https://keras.io.
[11] https://scikit-learn.org/stable/.
[12] https://nlp.stanford.edu/projects/glove/.

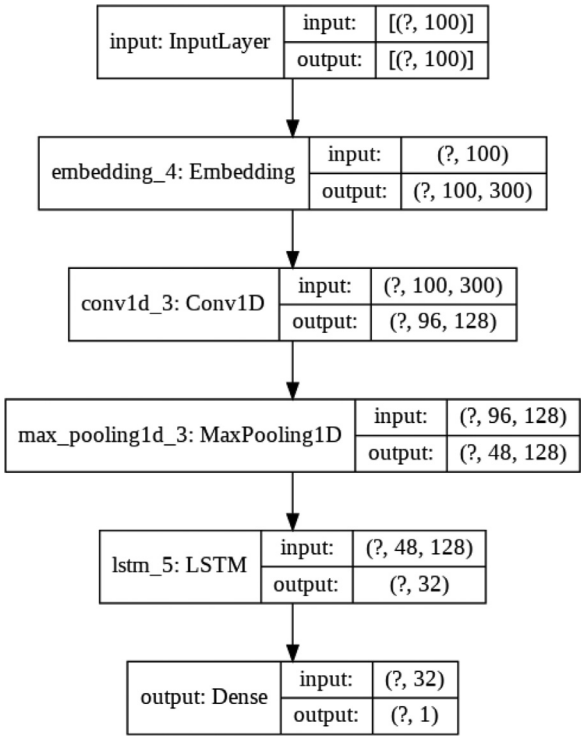

Fig. 5. ISOT Model Summary.

Table 4
Classification models used for comparison.

Sr. No.	Classifier	Shortname
1	Logistic Regression	LR
2	Random Forest	RF
3	Multinomial Naïve Bayes	MNB
4	Stochastic Gradient Decent	SGD
5	K Nearest Neighbors	KNNs
6	Decision Tree	DT
7	Ada Boost	AB
8	Convolutional Neural Networks	CNN only
9	Recurrent Neural Networks	RNN only

and n_o is the number of the LSTM output units (n_o=32) (Sak et al., 2014). The final dense layer connects the output of the LSTM with the output of the network (O=1), so the number of parameters is 32x1.

4.2.5. Comparison models

In order to comparatively evaluate the performance of the proposed model against other supervised classification methods, seven classifiers are tested (see Table 4), on both datasets using the same training-test split in all cases. The implementations of the Scikit-learn python library, with the default configuration has been used for all the comparison models. Results also comprise the performance of the proposed Neural Network approach when only the CNN or only the RNN layers are employed.

5. Evaluation and results

5.1. Evaluation metrics

For the evaluation of results, four metrics have been used, which are based on the number of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) in the predictions of the binary classifiers:

- Accuracy, which is the percentage of True (i.e. correct) predictions.
- Recall, which captures the ability of the classifier to find all the positive samples.
- Precision, which is the ability of the classifier not to label a negative sample positive.
- The F_1 score, which is the harmonic mean of precision and recall, computes values in the range [0,1].

The following equations compute the metrics:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$precision = \frac{TP}{TP + FP} \quad (2)$$

$$recall = \frac{TP}{TP + FN} \quad (3)$$

$$F_1\texttt{score} = \frac{2 * (precision \times recall}{precision + recall} \quad (4)$$

A paired t-test was used to validate the statistical significance of the results; the experiments were repeated five times (using 5-fold cross validation, i.e. 80%-20% split); and accuracy was reported at 95% confidence intervals.

Table 5
Results of all models on the FA-KES dataset.

Dataset	Accuracy	Precision	Recall	F_1score
LR	0.49 ± 0.008	0.50	0.49	0.49
RF	0.53 ± 0.009	0.56	0.53	0.54
MNB	0.38 ± 0.008	0.39	0.38	0.32
SGD	0.47 ± 0.009	0.49	0.47	0.48
KNNs	0.57 ± 0.008	0.58	0.57	0.57
DT	0.55 ± 0.006	0.56	0.55	0.55
AB	0.47 ± 0.005	0.49	0.47	0.47
(Elhadad et al., 2019)	0.58	0.63	0.58	0.50
CNN only	0.50 ± 0.006	0.55	0.50	0.48
RNN only	0.50 ± 0.007	0.51	0.50	0.50
Hybrid CNN-RNN	0.60 ± 0.007	0.59	0.60	0.59

Table 6
Results of all models on the ISOT dataset.

Dataset	Accuracy	Precision	Recall	F_1score
LR	0.52 ± 0.03	0.50	0.52	0.42
RF	0.92 ± 0.04	0.92	0.92	0.92
MNB	0.60 ± 0.02	0.60	0.60	0.60
SGD	0.52 ± 0.03	0.52	0.52	0.52
KNNs	0.60 ± 0.01	0.67	0.61	0.56
DT	0.96 ± 0.02	0.96	0.96	0.96
AB	0.92 ± 0.02	0.91	0.91	0.91
(Elhadad et al., 2019)	1.00	–	–	–
CNN only	0.99 ± 0.02	0.99	0.99	0.99
RNN only	0.98 ± 0.02	0.98	0.98	0.98
Hybrid CNN-RNN	0.99 ± 0.02	0.99	0.99	0.99

5.2. Results on the FA-KES dataset

The results of all methods on the FA-KES dataset are shown in Table 5. The table also shows the reported performance on the dataset as calculated by Elhadad et al. (2019).

The results show that the proposed Hybrid CNN-RNN method is significantly better than all other methods in terms of accuracy, precision, recall and F_1score.

5.3. Results on the ISOT dataset

The results of all methods on the ISOT dataset are shown in Table 6. The table also shows the reported performance on the dataset as calculated by Elhadad et al. (2019), who used uni-grams based SVM classifier for the task to obtain best results.

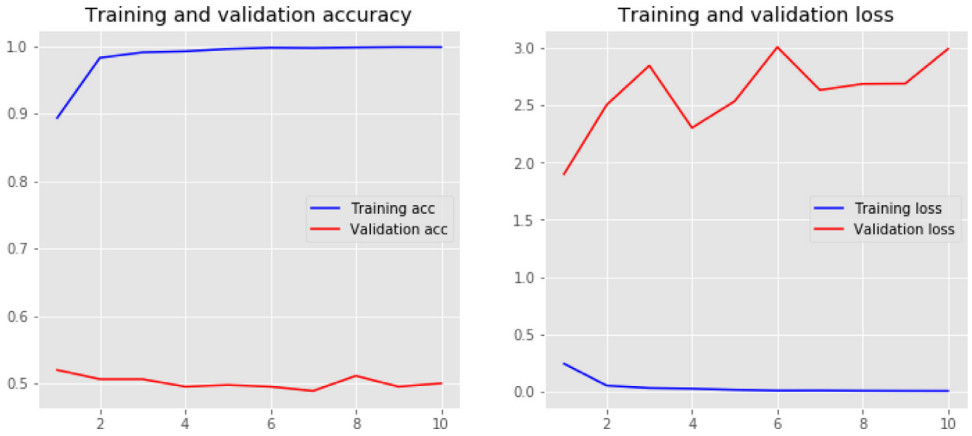

Fig. 6. Generalization training and validation accuracy and loss graphs.

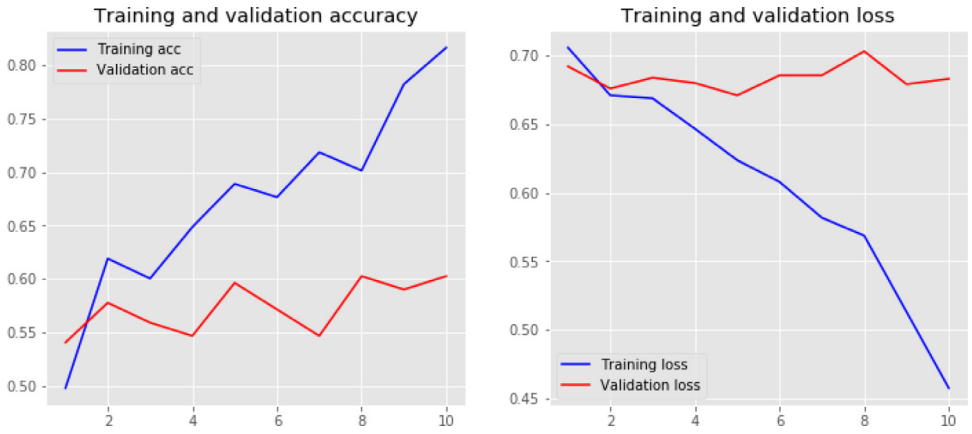

Fig. 7. FA-KES training and validation accuracy and loss graphs.

Although Elhadad et al. (2019) reported a 1.00 accuracy on the ISOT dataset with their Decision Tree classifier, the accuracy of their methods ranges between 0.85 and 1.00. It also seems that they used only part of the ISOT dataset, since they reported that only 25.2 thousand articles were used. Taking into account the statistical significance of the results (0.99 ± 0.02), the proposed method performs comparatively better than the state-of-the-art method of Elhadad et al. (2019).

Overall, none of the supervised classification methods performs better than the proposed hybrid CNN-RNN model. Among these methods, Random Forests performed better in terms of accuracy for both datasets. For FA-KES, KNNs had the best accuracy performance and for ISOT, it was Decision Trees.

6. Analysis of the results

Despite the many researches in the fake news detection domain, presented in Section 2, the issue of generalization of models still remains unnoticed. In order to promote the work in this domain, more experiments were performed with the hybrid CNN-RNN model trained on the ISOT dataset and tested on the FA-KES dataset using the exact same configuration as before. ISOT is chosen for training because it is much larger and has minimum space for improvement since many models perform above the 0.9 classification accuracy threshold. Fig. 6 shows the ability of the ISOT trained model to generalize on another dataset and plots the training and validation accuracy and loss values over the 10 epochs.

Results show that while the training accuracy and loss are optimum after 6 epochs, the validation accuracy remains almost the same in all epochs and is lower than that achieved when training (and validating) on the FA-KES dataset. The final performance after 10 epochs of training and test on the whole FA-KES dataset are shown in Table 7.

Table 7
Generalization results.

Accuracy	Precision	Recall	F1
0.50	0.48	0.48	0.46

The validation loss plot suggests an over-fitting of the model, since the loss fluctuates greatly with an increasing trend. The cross dataset validation results show poor generalisation, since despite the very large corpus used for training, and the near 1.0 accuracy of the model on training, it performs poorly on another fake news dataset using the exact same structure. Probably adding an internal drop-out layer could improve the ability of the model to generalise, and this will be further examined in the next work in the field.

In order to get a better understanding on whether the models are over-fitted to the data, the learning curves of the proposed method on the two datasets, the training and validation loss and accuracy curves over the 10 epochs on each dataset are plotted. The PyPlot module of the Matplotlib library has been used for this task and the results are depicted in Figs. 7 and 8.

For the ISOT dataset, the proposed model performed very well. The training and testing accuracy increases with the epochs and similarly, the respective loss values decrease, which indicates that the model learns to classify the articles better. For the FA-KES dataset, the training and validation accuracy do not hava a smooth increase (especially the validation accuracy). Similarly, the validation loss remains almost the same in all 10 epochs.

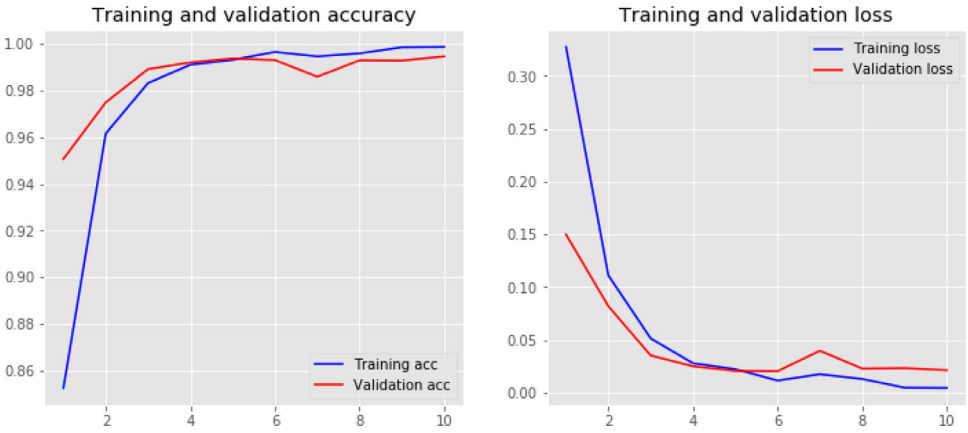

Fig. 8. ISOT training and validation accuracy and loss graphs.

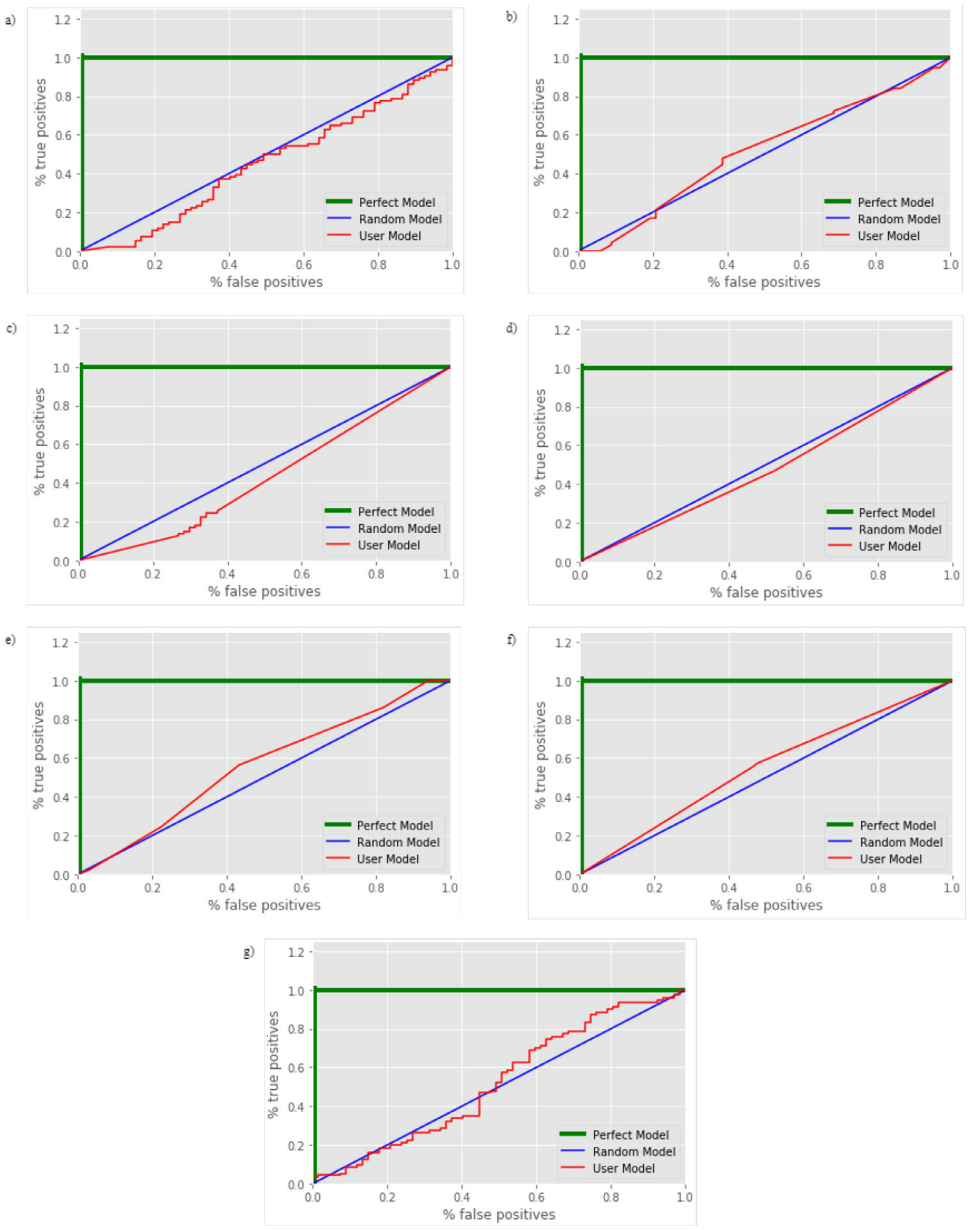

Fig. 9. FA-KES ROC curves. a) LR, b) RF, c) MNB, d) SGD, e) KNN, f) DT, g) AB.

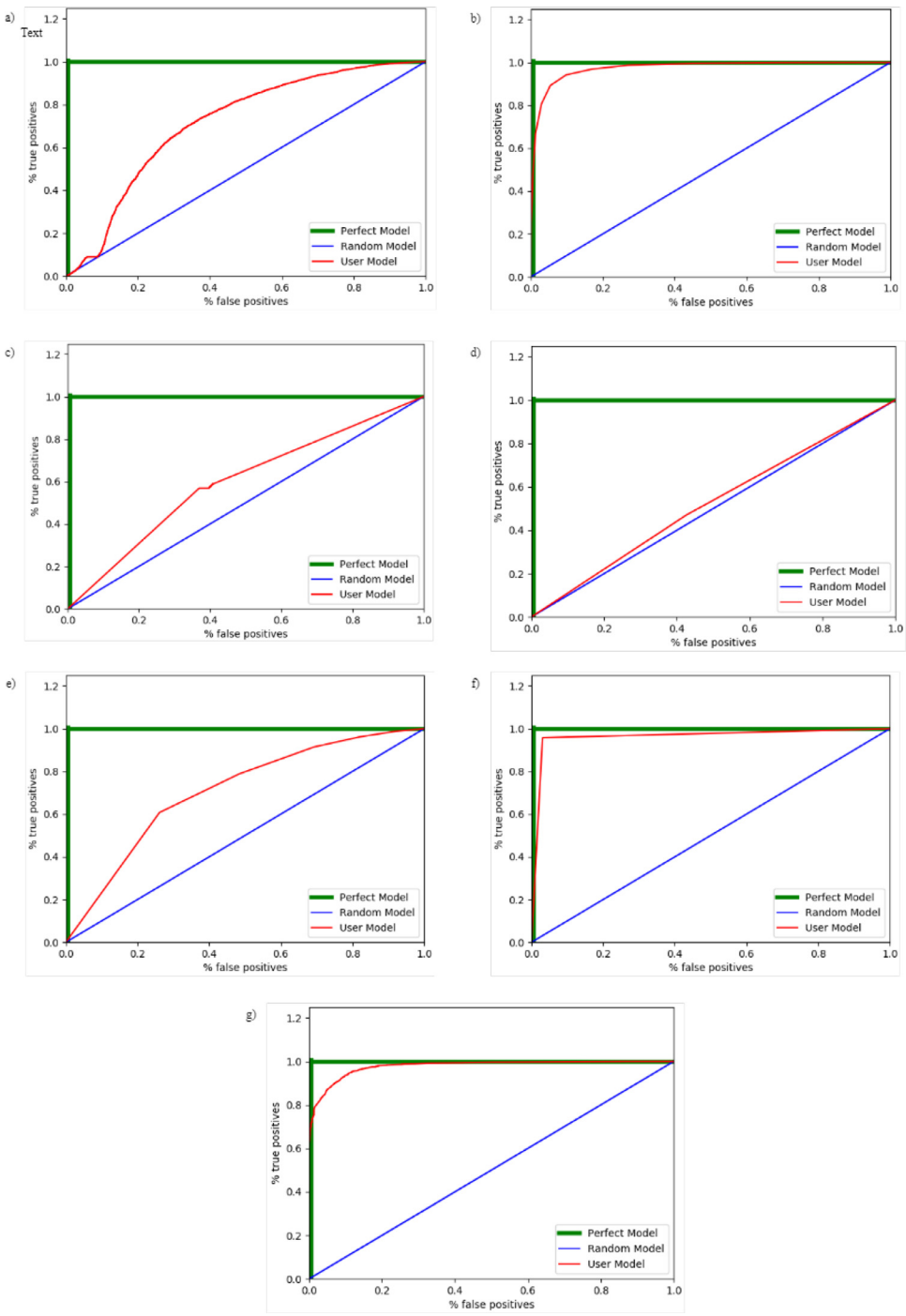

Fig. 10. ISOT ROC curves. a) LR, b) RF, c) MNB, d) SGD, e) KNN, f) DT, g) AB.

For completeness, the Receiver Operating Characteristic (ROC) curves for all the methods on the two datasets are plotted. The ROC curve is used to diagnose a binary classifier and is created by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). The ROC curves of the FA-KES dataset are depicted in Fig. 9 and the respective curves for ISOT are shown in Fig. 10.

The ROC curves of the AdaBoost (AB) and Random Forest (RF) algorithms are the most stable and the most accurate with a steady curve for ISOT. On the other hand, AB is the most unsteady curve for the FA-KES dataset. The models perform generally better on the ISOT dataset than on the FA-KES dataset.

7. Discussion

7.1. Contribution to theory

An important feature for predictive models that increases their applicability in multiple tasks, is their ability to generalise across datasets and tasks. *Model generalization* refers to the ability of a pre-trained model to handle unseen data and is mostly relevant to the model complexity and training. The majority of studies surveyed in Section 2 train deep neural networks on a dataset and evaluate their performance usually on a different subset of the same dataset. To the best of our knowledge,

there does not exist any study in the related literature of fake news detection that examines the ability of the models to generalize, and the current work is the first that contributes in this direction.

7.2. Implication to practice

A practical implication of training a model without examining its generalisation ability is *over-fitting*. The most obvious consequence of over-fitting is the poor model performance on a new, unseen dataset. In addition, over-fitted models expose high-complexity, and examine a lot more information than it is probably needed in order to reach a decision. Finally, over-fitted models cannot be ported to a similar task on another dataset and require re-training from scratch, thus reducing reusability.

Another practical implication of deep neural networks relates to their specialisation in certain tasks, which reduces their ability to perform well in different tasks. The restriction of neural networks to solve on problem at a time, becomes the advantage of the proposed hybrid method, which combines a CNN network that learns the spatial, thus conceptual, features of text, with an LSTM that captures the sequential flow of text. The main hypothesis that a hybrid Convolutional Neural Networks (CNN) Recurrent Neural Networks (RNN) model could improve on state-of-the-art baselines for fake news detection is experimentally confirmed. Experiments on two real-world fake news datasets of different nature (∼100% accuracy on ISOT dataset, 45000 articles; ∼60% accuracy on FA-KES dataset, 804 articles) demonstrate significantly better results than the non-hybrid baseline methods.

Finally, deep learning algorithms can be benefited from the abundance of training data. This study has surveyed many human annotated datasets that can be used for training a model that successfully detects fake news. A generalised model that is able to capture the features of each dataset and learn what is valid and what is not, can be trained on multiple datasets, given that they are properly formatted and fed to the model.

8. Conclusion

Despite the relative abundance of extant works addressing fake news detection, there is still plenty of space for experimentation, and the discovery of new insights on the nature of fake news may lead to more efficient and accurate models. In addition, this paper is, to the best of our knowledge, the first to suggest generalization of models used for fake news detection. As shown in this work, such models tend to work well on a specific dataset, but do not generalize well. New horizons can be explored by considering generalization of fake news detection models.

Overall, the use of artificial neural networks seems promising in fake news detection. Apart from CNN and RNN, more complex neural network architectures will be considered as part of our future analysis. Traditional models can also be beneficial if they are combined with task-specific feature engineering techniques.

References

Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. *Security and Privacy, 1*(1), e9.

Ajao, O., Bhowmik, D., & Zargari, S. (2018). Fake news identification on twitter with hybrid CNN and RNN models. In *Proceedings of the 9th international conference on social media and society* (pp. 226–230).

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics, 6*(2). 2053168019848554

Aswani, R., Kar, A. K., & Ilavarasan, P. V. (2018). Detection of spammers in twitter marketing: a hybrid approach using social media analytics and bio inspired computing. *Information Systems Frontiers, 20*(3), 515–530.

Aswani, R., Kar, A. K., & Ilavarasan, P. V. (2019). Experience: Managing misinformation in social media-insights for policymakers from twitter analytics. *Journal of Data and Information Quality (JDIQ), 12*(1), 1–18.

Bouchra, N., Aouatif, A., Mohammed, N., & Nabil, H. (2019). Deep belief network and auto-encoder for face classification. *IJIMAI, 5*(5), 22–29.

Buntain, C., & Golbeck, J. (2017). Automatically identifying fake news in popular twitter threads. In *2017 IEEE international conference on smart cloud (smartcloud)* (pp. 208–215). IEEE.

Derczynski, L., Bontcheva, K., Liakata, M., Procter, R., Hoi, G. W. S., & Zubiaga, A. (2017). Semeval-2017 task 8: Rumoureval: Determining rumour veracity and support for rumours. arXiv:1704.05972.

Derczynski, L., Bontcheva, K., Lukasik, M., Declerck, T., Scharl, A., Georgiev, G., … Stewart, R., et al. (2015). Pheme: Computing veracity-the fourth challenge of big social data. In *Proceedings of the extended semantic web conference eu project networking session (ESCW-PN)* (p. n/a).

Dhillon, A., & Verma, G. K. (2020). Convolutional neural network: A review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence, 9*(2), 85–112.

Drumond, T. F., Viéville, T., & Alexandre, F. (2019). Bio-inspired analysis of deep learning on not-so-big data using data-prototypes. *Frontiers in Computational Neuroscience, 12*, 100.

Elhadad, M. K., Li, K. F., & Gebali, F. (2019). A novel approach for selecting hybrid features from online news textual metadata for fake news detection. In *International conference on p2p, parallel, grid, cloud and internet computing* (pp. 914–925). Springer.

Ferreira, W., & Vlachos, A. (2016). Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 1163–1168).

Gorrell, G., Kochkina, E., Liakata, M., Aker, A., Zubiaga, A., Bontcheva, K., & Derczynski, L. (2019). Semeval-2019 task 7: Rumoureval, determining rumour veracity and support for rumours. In *Proceedings of the 13th international workshop on semantic evaluation* (pp. 845–854).

Granger, E., Kiran, M., Blais-Morin, L.-A., et al. (2017). A comparison of CNN-based face and head detectors for real-time video surveillance applications. In *2017 seventh international conference on image processing theory, tools and applications (IPTA)* (pp. 1–7). IEEE.

Graves, A. (2012). Long short-term memory. In *Supervised sequence labelling with recurrent neural networks* (pp. 37–45). Springer.

Hamdi, T., Slimi, H., Bounhas, I., & Slimani, Y. (2020). A hybrid approach for fake news detection in twitter based on user features and graph embedding. In *International conference on distributed computing and internet technology* (pp. 266–280). Springer.

Jin, Z., Cao, J., Zhang, Y., & Luo, J. (2016). News verification by exploiting conflicting social viewpoints in microblogs. In *Thirtieth AAAI conference on artificial intelligence* (pp. 2972–2978).

Kar, A. K. (2016). Bio inspired computing–a review of algorithms and scope of applications. *Expert Systems with Applications, 59*, 20–32.

Karimi, H., & Tang, J. (2019). Learning hierarchical discourse-level structure for fake news detection. arXiv:1903.07389.

Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political astroturfing on twitter: How to coordinate a disinformation campaign. *Political Communication, 37*(2), 256–280.

Khan, J. Y., Khondaker, M., Islam, T., Iqbal, A., & Afroz, S. (2019). A benchmark study on machine learning methods for fake news detection. arXiv:1905.04749.

Kollias, D., & Zafeiriou, S. P. (2020). Exploiting multi-CNN features in CNN-RNN based dimensional emotion recognition on the OMG in-the-wild dataset. *IEEE Transactions on Affective Computing*.

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

Kumar, S., West, R., & Leskovec, J. (2016). Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In *Proceedings of the 25th international conference on world wide web* (pp. 591–602).

Li, Y., Gao, J., Meng, C., Li, Q., Su, L., Zhao, B., … Han, J. (2016). A survey on truth discovery. *ACM SIGKDD Explorations Newsletter, 17*(2), 1–16.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing, 234*, 11–26.

Long, Y. (2017). *Fake news detection through multi-perspective speaker profiles*. Association for Computational Linguistics.

Lukasik, M., Cohn, T., & Bontcheva, K. (2015). Classifying tweet level judgements of rumours in social media. arXiv:1506.00468.

Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K.-F., & Cha, M. (2016). Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the twenty–fifth international joint conference on artificial intelligence (IJCAI-16)* (pp. 3818–3824). AAAI Press.

Ma, J., Gao, W., Wei, Z., Lu, Y., & Wong, K.-F. (2015). Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the 24th ACM international on conference on information and knowledge management* (pp. 1751–1754).

Masood, S., Srivastava, A., Thuwal, H. C., & Ahmad, M. (2018). Real-time sign language gesture (word) recognition from video sequences using CNN and RNN. In *Intelligent engineering informatics* (pp. 623–632). Springer.

Meel, P., & Vishwakarma, D. K. (2019). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, 112986.

Mihaylova, T., Karadjov, G., Atanasova, P., Baly, R., Mohtarami, M., & Nakov, P. (2019). Semeval-2019 task 8: Fact checking in community question answering forums. arXiv:1906.01727.

Mitra, T., & Gilbert, E. (2015). Credbank: A large-scale social media corpus with associated credibility annotations.. In *ICWSM* (pp. 258–267).

Pamungkas, E. W., Basile, V., & Patti, V. (2019). Stance classification for rumour analysis in twitter: Exploiting affective information and conversation structure. arXiv:1901.01911.

Pierri, F., & Ceri, S. (2019). False news on social media: a data-driven survey. *ACM SIGMOD Record, 48*(2), 18–27.

Popat, K., Mukherjee, S., Yates, A., & Weikum, G. (2018). Declare: Debunking fake news and false claims using evidence-aware deep learning. arXiv:1809.06416.

Rajagopal, A., Joshi, G. P., Ramachandran, A., Subhalakshmi, R., Khari, M., Jha, S., Shankar, K., & You, J. (2020). A deep learning model based on multi-objective particle

swarm optimization for scene classification in unmanned aerial vehicles. *IEEE Access, 8*, 135383–135393.

Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 2931–2937).

Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011). Truthy: Mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th international conference companion on world wide web* (pp. 249–252).

Ruchansky, N., Seo, S., & Liu, Y. (2017). Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on conference on information and knowledge management* (pp. 797–806).

Sahoo, S. R., Gupta, B., Choi, C., & Esposito, C. (2020). Detection of spammer account through rumor analysis in online social networks. In *The 9th international conference on smart media and applications* (p. n/a).

Sak, H., Senior, A., & Beaufays, F. (2014). Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. arXiv:1402.1128.

Salem, F. K. A., Al Feel, R., Elbassuoni, S., Jaber, M., & Farah, M. (2019). Fa-kes: A fake news dataset around the syrian war. In *Proceedings of the international aaai conference on web and social media: 13* (pp. 573–582).

Shang, J., Shen, J., Sun, T., Liu, X., Gruenheid, A., Korn, F., … Han, J. (2018). Investigating rumor news using agreement-aware search. In *Proceedings of the 27th ACM international conference on information and knowledge management* (pp. 2117–2125).

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. arXiv:1809.01286, 8.

Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). Fever: a large-scale dataset for fact extraction and verification. arXiv:1803.05355.

Vlachos, A., & Riedel, S. (2014). Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 workshop on language technologies and computational social science* (pp. 18–22).

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151 .

Wang, W. Y. (2017). " liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv:1705.00648.

Wang, Y., McKee, M., Torbica, A., & Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. *Social Science & Medicine, 240*, 112552.

Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019). Misinformation in social media: Definition, manipulation, and detection. *ACM SIGKDD Explorations Newsletter, 21*(2), 80–90.

Yadav, A., & Vishwakarma, D. K. (2020). A comparative study on bio-inspired algorithms for sentiment analysis. *Cluster Computing*, 1–21.

Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2018). Ti-CNN: Convolutional neural networks for fake news detection. arXiv:1806.00749.

Zeng, L., Starbird, K., & Spiro, E. S. (2016). #unconfirmed: Classifying rumor stance in crisis-related social media messages. In *Tenth international AAAI conference on web and social media* (pp. 747–750).

Zhang, X., Chen, F., & Huang, R. (2018). A combination of rnn and cnn for attention-based relation classification. *Procedia Computer Science, 131*, 911–917.

Zhou, C., Sun, C., Liu, Z., & Lau, F. (2015). A c-lstm neural network for text classification. arXiv:1511.08630.

Zhou, X., Jain, A., Phoha, V. V., & Zafarani, R. (2020). Fake news early detection: A theory-driven model. *Digital Threats: Research and Practice, 1*(2), 1–25.

Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR), 51*(2), 1–36.

Zubiaga, A., & Ji, H. (2014). Tweet, but verify: epistemic study of information verification on twitter. *Social Network Analysis and Mining, 4*(1), 163.

Zubiaga, A., Kochkina, E., Liakata, M., Procter, R., & Lukasik, M. (2016). Stance classification in rumours as a sequential task exploiting the tree structure of social media conversations. arXiv:1609.09028.

Zubiaga, A., Liakata, M., & Procter, R. (2017). Exploiting context for rumour detection in social media. In *International conference on social informatics* (pp. 109–123). Springer.