Journal of Theoretical and Applied Information Technology
15th March 2023. Vol.101. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ENHANCING ARABIC FAKE NEWS DETECTION FOR TWITTERS SOCIAL MEDIA PLATFORM USING SHALLOW LEARNING TECHNIQUES

ALBARA AWAJAN
Intelligent Systems Department, Al-Balqa Applied University
Al-Salt, Jordan
E-mail: a.awajan@bau.edu.jo

ABSTRACT

One of the most significant and severe issues facing social media nowadays is fake news, especially in social media applications - the intentional deception of users through the spread of propaganda, rumors, or misleading information about a variety of people or issues. A very common social media platform with the highest rapidly increasing number of users in the Middle East is Twitter. However, along with an increase in the number of Twitter users has come an increase in the amount fake news being spread across the platform. This, in turn, has caught the attention of researchers who seek nothing more than an online experience free of fake news. As are result, through utilizing both the transformer-based language and recurrent neural network models, this paper introduces an intelligent classification model that identifies fake news presented in Arabic tweets. Afterwards, this research presents a comparative study between deep learning and shallow learning. In Addition, to enhance the effectiveness of the proposed model, researchers also built an Arabic Twitter dataset that included 206,080 tweets. Results have shown that pre-trained deep learning models performed much better than shallow models when it came to identifying fake news in Arabic. That is, a 95.92% accuracy rate was reached when researchers applied shallow learning to the pre-processed data set; however, after applying the LSTM model to the same data, the accuracy rate increased to 96.71% while after apply the Bert model to that same data, the accuracy rate increased to a near perfect 99%.

Keywords: *Fake news, Shallow Learning, Deep learning, Twitter, Social Media.*

1. INTRODUCTION

A very significant research topic among the Natural Language Processing (NLP) and Machine Learning (ML) communities is that of text classification or organization. This is due to many factors including the huge global access to text data on the world wide web [1]. What is more, it can be used in many applications including Topic Classification [2], Information Retrieval [3], Sentiment Analysis [4], Email Filtering and Spam Detection [5], Fake News Classification [6], and Word Disambiguation [7].

Nowadays, social media networks are a primary source for gathering information and receiving news for Middle Easterners. Indeed, the explosion of social media applications and the number of people using them has allowed users to share information with little to no research or filtering being applied – all for free. Add to this the statistics from Radcliffe and Bruni which show that social media is the first source they gather news from for about 63% of Arab youths [1], and one can easily conclude how this has not only compounded the issue of fake news, but also created major problems in Arab society today due to its negative effects.

The definition of fake news is false and often sensational news stories created to be shared or distributed far and wide for the purpose of achieving a specific goal such as making money or supporting or harming the reputation of a public figure, political movement, company, and so on. Keeping these false news stories in check has, indeed, become a huge challenge for social networks like Facebook and Twitter. Additionally, it has devastating social, political and economic effects on countries and communities worldwide. For instance:

- In April 2013, a fake tweet was spread via Twitter about two explosions at the White House which resulted in a public scare

and a big decrease in stock market prices [2].

- Fake news about a Malaysian airplane being lost in Weibo in March 2014 brought fear and panic to the passenger's families which made people confused and not able to follow the factual news [3].
- During the 2016 United States presidential election, around 530 different fake news stories about the presidential candidates were disseminated via Twitter and Facebook, all of which had some effect on voters [4].

A major step in solving a problem in a machine learning society is choosing the most apt classifier, while in text classification, feature representation based on the bag-of-words (BoW) model is usually used to mine the unigram or n-gram as the component features. The main issue with that, however, is that, because of the properties of Arabic words like synonymy and polysemy make Arabic such an intricate language, utilizing the n-grams model on Arabic text classification is not always successful [5]. In fact, this was one of the issues that inspired a number of different attempts at find more complex methods of extracting features and classifier schemes that would provide more appropriate representations of the content in a many text classification task including that which can be classified as fake news. As a result, in 2012, a widely accepted solution was introduced in 2012 – deep learning [6].

The main contributions presented in this work are:

- to create a new Arabic dataset using Twitter API.
- to offer an effective model that detects Arabic fake news text utilizing LSTM and BERT deep learning models.
- to compare our model with a variety of state-of-the-art machine learning models

The work presented in the paper aims to address the growing spread of fake news on social media platforms. It also focuses on the Arabic language because it is one of the most widespread spoken languages in the world, but it has very little interest in NLP in general and Fake news detection in particular. The paper also implements and compares shallow learning and deep learning to ensure that the final model meets the highest expectations. In order to achieve this goal, a well-defined and appropriate dataset based on Arabic Twitter tweets must be gathered and built.

The paper from now on is organized as follows: In Section 2, related literature is explored whereas Section 3 the methodology used in carrying out the studies is reviewed and the various deep learning algorithms that were used in creating the classification model are outlined. Classification experiments and their results are described in Section 4, and finally, the study presents its conclusions in Section 5.

2. RELATED WORKS

Much research has been carried out of late which has addressed the issue of classifying fake news, disinformation, rumors or misinformation found on social media. Most of this research can be divided into two approaches – shallow learning and deep learning – depending on the proposed classification systems. This section gives an overview of already existing techniques and approaches used in detecting and classifying fake news.

2.1 Arabic Fake News

While there have been many studies carried out on detecting and classifying fake news in the English language, very few studies have been done on the subject of detecting and classifying fake news in Arabic.

2.1.1 Shallow Learning

Two studies carried out by Mahlous and Al-Laith [7] and Thaher, et al. [8] presented a machine learning model to detect fake news in Arabic spread via Twitter.

Mahlous and Al-Laith [7] focused their research on issues with detecting Arabic fake news tweets about COVID-19. Utilizing Twitter's API and Tweepy Python library, they gathered around seven million Arabic tweets related to COVID-19. After deleting repeated tweets, they applied a variety of pre-processing techniques which included removing hyperlinks, hashtags, mentions, strange and non-Arabic words, Arabic diacritics and the repeated characters, they managed to filter the amount of Arabic tweets down to about 5.5 million. Furthermore, by manually annotating a sample of 2500 tweets as either fake or non-fake, the authors also, created a fake news annotation system. They also prepared the attributes needed to create classification models by utilizing feature extraction methods including Count Vector, N-gram-Level TF-IDF, Character-Level TF-IDF and

Word-Level TF-IDF. Additionally, they also used six machine learning classifiers: Naïve Bayes, Logistic Regression, Multilayer Perceptron, the Support Vector Machine, the eXtreme Gradient Boosting Model and the Random Forest Bagging Model. Using Logistic Regression, Mahlous and Al-Laith's model gained an F1 core of 93.3%.

On the other hand, Thaher, et al. [8] put forward an intelligent detection system for fake Arabic Twitter corpus news using Natural Language Processing (NLP) methods. The researchers utilized eight machine learning algorithms and different feature extraction models to determine the best suitable models that can be used for identifying Arabic fake news tweets. Their model consists of six phases. First, Thaher, et al. utilized the Syrian Crisis tweets dataset first collected by Al Zaatari, et al. in their study [9]. Next, they cleaned that dataset from irrelevant text, then utilized normalization, tokenization, stemming, and text vectorization techniques to filter texts down even more. After this, they then extracted features from the filtered dataset using TF, BTF and TF-IDF in order to get structured numerical features from the unstructured tweet text. This was followed by using eight machine learning classifiers - NB, KNN, SVM, DT, LR, Linear Discriminant Analysis (LDA), Random Forest (RF), and XGBoost – testing each of them within similar conditions using the Python Scikit-learn ML library. Finally, they utilized the Binary Harris Hawks Optimizer (HHO) as a wrapper-base feature selection method. The experiments concluded that when applying the LR classifier with TF-IDF model ranked best with an accuracy rate of 81.5%.

Recent research carried out by Alkhair, et al. [10] employed YouTube replies and comments to investigate the fake news content in the Middle East where the data was gathered by posts on the YouTube application. The contribution of Alkhair and his team can be summed up in three points: First, by collecting 4079 comments about three selected famous celebrities in the Arab world using YouTube API, and then removing the special characters, words in foreign languages, URL links and repeated comments to clean the data, they introduced a new Arab corpus for fake news research and analysis. Second, they provided researchers with a variety of analyses on what data researchers need to retrieve when looking for information about the subject of fake news detection. Finally, Alkhair, et al. used three machine learning classifiers: Multinomial Naïve Bayes (MNB), Support Vector Machine (SVM) and Decision Tree (DT) to test probability between rumors and facts of the YouTube comments. In the end, using the SVM classifier enabled them to obtain a 95.35% accuracy rate.

Most recently, using text mining techniques, Sarah Alanazi [11] relied on analyzing people's comments after reading the news on social media to propose a novel fake news detection system. Two environments were used by the researchers to perform these experiments - RapidMiner and Python. First, they created an Arabic news dataset and then applied pre-processing techniques such as Tokenization, Removal of Stop Words and Stemming to the data utilizing four machine learning classifiers which included Decision Tree (DT), Naïve Bayes (NB), Support Vector Machine (SVM) and K-NN. As a result, their experiments proved that the NB classifier outperformed other classifiers by attaining 87.18% accuracy in the RapidMiner environment. Likewise, the SVM classifier was the most promising classifier in the Python environment achieving an accuracy rating of 87.14%.

2.1.2 Deep Learning

Considering deep learning methods, Nagoudi, et al. [12], Al Yahya, et al. [13] and Harrag, et al. [14] used the transformer-based language model (AraBERT) to create a fake news detection model; whereas, Saadany, Mohammad, and Orasan [15] used a deep learning model with a proposed combination of a CNN model and a pre-trained word embeddings model.

Nagoudi, et al. [12] resolved the issue of there not being enough suitable data to train detection models with by suggesting a method for automatically generating potentially fake Arabic news stories. Further, they also created models to identify Arabic news that had been manipulated which resulted in successfully finding fake news in Arabic. Utilizing both ATB and their introduced AraNews datasets, they evaluated their model. Splitting the data into three groups - 80% for training, 10% for development and 10% for testing – they utilized the Khouja dataset which includes 3,072 true sentences and 1,475 fake sentences to assess their model on a human generated fake news dataset. Moreover, they also used four pre-trained masked language models (MLM): mBERT, AraBERT, XLM-RBase, and XLM-RLarge, reaching an accuracy rating of 89.23% with the AraBERT pre-trained model.

On the other hand, Al Yahya, et al. [13] proposed a model that compared deep learning

models to transformer-based models through a comparative study of neural network and transformer-based language models presenting and comparing the performance of each of these models when detecting fake news in Arabic. Using CNN, RNN, and GRU as their deep learning models and ArElectra, QARiB, Marbert, Arbert (AraBERT v1, AraBERT v02 and AraBERT v2) as their transformer-based models, Al Yahya, et al. designed a number of experiments. Furthermore, they utilized COVID-19-Fakes [16], AraNews [12], ArCOV19-Rumors [17] and ANS corpus [18] datasets that were divided into training set and a validation set where the ratio was 80 to 20 realtivly. After this, they applied the pipelines for text preprocessing phase. The results show that the transformer-based models perform much better than the neural network-based solutions perform, with the GRU model reaching the highest accuracy rating at 95%.

Another example is the comparison made by Harrag, et al. [14] between the LSTM model and the AraBERT model. In order to determine whether sentences in Arabic written by bots were fake or not, they suggested using a transfer learning-based model, expanding upon the dataset proposed by Almerekhi and Elsayed in their 2015 study Detecting Automatically-Generated Arabic Tweets in AIRS [19] and by introducing a new set of 4,196 tweets for detecting deepfake texts that they divided into 80% for training and 20% for testing . For evaluation, Harrag, et al. used a variety of recurrent neural network (RNN) word embeddings-based models including LSTM, BI-LSTM, GRU and BI-GRU, comparing them to AraBERT (BERT model for Arabic). As a result they obtained an accuracy rating of 98.7% using their model along with the AraBERT model.

By performing several analyses to determine the linguistic properties of fake news in Arabic, Saadany, Mohammad, and Orasan [15] also proposed a detection system based on distinguishing features at the lexico-grammatical level. Their method is made up of an amalgamation of a CNN with pre-trained word embeddings with no hand-crafted linguistic features. The group of researchers built a new dataset by gathering 3,185 pieces of fake news crawled from two sites: Al-Hudood and Al-Ahram Al-Mexici. They added to these 3,710 pieces of real news they crawled from the official news sites BBC-Arabic, CNN-Arabic and Al-Jazeera. Three classification models were used: 1.) Naive Bayes classifier as a baseline model with both TF-IDF and 9Bag-Of-Words (BOW) approaches as numerical features, ultimately obtaining an accuracy rating of 96.23%. 2.) XGBoost with Count Vectors that obtained an accuracy rating of 96.81%. 3.) CNN with pre-trained word embedding that, at 98.59%, achieved the highest accuracy rating among the three models.

2.2 English Fake News

In contrast to Arabic language context, a large amount of literature has been applied to the classification of English language fake news.

2.2.1 Shallow Learning

Promoting their new set of features for training classifiers, Reis, et al. [20] measured the prediction of current approaches and features performance for automatic fake news detection. By utilizing a dataset containing 2,282 BuzzFeed news articles, they realized more than any other features put forth and explored in previous works. Pre-processing data by removing stories labeled as illegitimate content, they blended news that was mostly untrue with a combination of true and untrue news stories into a single fake news classification. In most cases, certain factors for fake news detection can be removed from news content, news sources and environment. Reis, et al. used Naive Bayes (NB), k-Nearest Neighbors (KNN), Random Forests (RF), XGBoost (XGB), Support Vector Machine (SVM) to assess the discriminative power of the previous features. In the end, using RF and XGBoost classifiers attained the highest accuracy rating at 86%.

Similar contributions have been made by other researchers such as Ahmed, Traore, and Saad [21] who used n-gram analysis and machine learning techniques to create a fake news detection model. Their model is made up of five main phases. First, they create a new dataset by gathering 12,600 legitimate news articles from reuters.com along with another 12,600 fake news articles from a fake news dataset on kaggle.com . In addition to these datasets, they used the Horne and Adali dataset for validation of their model. Second, Ahmed, Traore, and Saad used two pre-processing techniques - Stop Word Removal and Stemming – to clean the dataset. Third, to create features from the documents, they then utilized the word-based n-gram feature extraction method. Fourth, to remove the irrelevant and redundant features, Ahmed, Traore, and Saad next used two features selection techniques - The Term Frequency (TF) and the Term Frequency-Inverted Document Frequency (TF-IDF). Finally, using six different machine classification techniques - Support Vector Machines (SVM), Linear Support Vector Machines

(LSVM), Stochastic Gradient Descent (SGD), K-Nearest Neighbor (KNN) and Decision Trees (DT) to predict the class of the documents, the researchers trained and tested their model. They split the dataset to 80% for training and 20% for testing with 5-fold cross validation. Using the TF-IDF as a feature extraction and LSVM as a classifier, their model was able to attain an accuracy rating of 92%.

Likewise, to identify fake news on social media, Ozbay and Alatas [22] offered a two-step method: Step one applied a number of pre-processing techniques like Tokenization, Removing Stop Words and Stemming to convert unstructured data to a more structured data within the dataset. They then calculated the Term Frequency (TF) and created Document Term Matrix methods to represent the data through vectors. Step two employed 23 supervised machine learning algorithms. Ozbay and Alatas used three datasets to assess the supervised algorithms - the BuzzFeed Political News Data set [23] which included 1,627 news articles, the Random Political News Data set containing 75 news articles, and the ISOT Fake News Data set that included 44,898 articles. Utilizing the Decision Tree classifier on the ISOT Fake News data set yielded the highest accuracy rating at 96.8%.

One final example approach was used by Jain, et al. [24] which presented a new methodology for fake news detection. There were three main models that made up their proposed model: the News Aggregator which allows users to view news and information from a variety of sources all in one location; the News Authenticator which helps determine whether the news is legitimate or illegitimate by comparing the news on their side with different authored websites; and the News Suggestion System which recommends correlated news if the user had presented fake news to the system. Additionally, Jain, et al. also used the Support Vector Machine (SVM) and the f Naïve Bayes (NB) machine learning classifiers to test their method with their proposed system accuracy rating reaching 93.6%.

2.2.2 Deep Learning

Veyseh, et al. [25], Ghanem, et al. [26], and Zhou, Wu, and Zafarani [27] suggested new fake news detection models, while on the other hand, Wani, et al. [28] compared different state-of-the-art deep learning classifiers like LSTM, CNN, and BERT models. Moreover, a combination of complex and recurrent neural networks was proposed by Nasir, Khan, and Varlamis [29] to solve the detection problem.

To detect the rumors that spread in Twitter, Veyseh, et al. [25] proposed a semantic graph model. To validate their system Veyseh, et al. used the Twitter15 and Twitter16 datasets containing 1,381 tweets and 1,118 tweets, respectively. They used two models – featured based models like Decision Tree, SVM, and Random Forest and deep learning models like GRU-RNN, BU-RvNN, and TD-RvNN – to compare their model to other state-of-the-art methods. Their best accuracy rating was 76.8% among all other models compared. However, perhaps the biggest contribution to come out of this research by Veyseh, et al. is a system that was taught how implied relations between a tweet and its replies is based on the content of tweets rather than the direct reply relationship.

More recent research conducted by Ghanem, et al. [26] proposed the FakeFlow system to detect the fake news articles. By replicating the flow of a neural architecture, they assessed the model's performance using MultiSourceFake, a dataset they created themselves which includes 11,397 articles along with another three different available datasets: TruthShades which includes 23,000 articles [30], PoliticalNews which includes 14,240 articles [31] and FakeNewsNet which includes 20,208 articles [32]. To evaluate the FakeFlow model, Ghanem, et al. used a combination of fake news detection models (FakeNewsDetector [33], Horne and Adali [23], Rashkin [30], EIN [34]) and deep neural network architectures (CNN, LSTM, BERT, HAN [35],and Longformer [36]). Compared to other state-of the-art methods, their system obtained a much more superior accuracy rating of 96%.

Moreover, Zhou, Wu, and Zafarani [27] put forward a multi-modal (textual and visual) model they called the Similarity-Aware Fake news detection model (SAFE) to examine the role that the similarities between visual and textual news content plays in fake news detection. It consists of three main modules: 1.) multi-modal feature extraction, 2.) modal independent fake news prediction, and 3.) cross-modal similarity extraction. Additionally, they espoused Text-CNN to extract textual and visual features from the articles for news representation. They also used two datasets to carry out their experiments: PolitiFact (which includes 1,056 news articles) and GossipCop (containing 22,140 news articles). After comparing their model with state-of-the-art fake news detection models that included textual LIWC

[37], visual VGG19 [38] and multi-modal information att-RNN [39], Zhou, Wu, and Zafarani's system attained an accuracy rating of 87.4%.

Still, Nasir, Khan, and Varlamis [29] used another approach that put forth a hybrid deep learning model which includes a amalgamation of complex and recurrent neural networks to solve the problem of fake news classification. To extract the local features they used a CNN layer of Conv1D, processing the input vectors. Next, to learn long-term dependencies of the local features to classify the news articles to real or fake, the CNN layer output was used as the input for the RNN layer of the LSTM unit. Nasir, Khan, and Varlamis validated their system using two datasets: FA-KES (which contains 804 news articles) and ISOT (which contains 45,000 news articles). Using the ISOT dataset, Nasir, Khan, and Varlamis' hybrid model attained a 99% accuracy rating.

In sum, the researched literature reviews validated the effectiveness of deep learning and supervised machine learning approaches for detecting and classifying fake news in tweets. Nevertheless, there is a still a shortage of fake news datasets, and particularly in Arabic. Thus, the archaic Arabic fake news classification of fake news in social media platforms requires further study. Additionally, as a result of high dimensionality of feature space in the text classification domain, there will be an increase in the performance of the classification model due to noise and other irrelevant features may be present. These facts encouraged us to propose a more proficient detection model for the classification of Arabic fake news on the Twitter platform through adapting recent deep learning approaches such as LSTM and BERT as feature selection and classification models. A comparision between fake news detection using shallow learning and deep learning models is shown in table 1.

After a thorough review of the literature and a thorough comparison of the models used in each, it is clear that it is essential to develop a specialized dataset for Arabic fake news for Twitter in order to successfully achieve an accurate system with superior performance. After constructing a reliable dataset, various types of machine learning, deep learning, and shallow learning algorithms must be applied to the dataset in order to produce an effective model that detects Arabic fake news with the highest performance. The implemented model should also be compared to a variety of advanced machine learning models to ensure accuracy and performance in comparison to these models.

3. METHODOLOGY

In this section we will present the methodology for both shallow learning approach and deep learning approach

3.1 Shallow Learning Approach

A sub-field of Artificial Intelligence (AI), shallow learning relies on humans to determine which set of features will be utilized to assess differences between input data. Shallow learning is comprised of three central parts: 1.) the decision process to create assessments about data patterns as a result of input; 2.) an error function to assess model prediction; and 3.) a model optimization process which adjusts the weights to lessen conflict between the model estimate and the known example. There has been an increase in shallow learning in many areas such as mobile malware detection [40-45], malware and intrusion detection [46-50], and information retrieval [51]. To train the algorithm utilizing a labeled dataset to identify fake news in Arabic tweets, a supervised learning model was implemented in this research. Figure 1 shows the model that was implemented:

3.2 Deep Learning Approach

One of the most popular machine learning subfields to appear recently is deep learning. Basically, deep learning is a neural network (NN) with more than two layers of interconnected nodes, with each one building upon the prior layer to perfect the prediction or the classification mechanisms. Deep learning techniques have gained much attention lately in comparison to classical machine learning methods due to their advanced precision and speed of cluster data in making predictions [52].

3.2.1 Siamese LSTM Architecture

As proposed in Tan, Wang and Lee's 2002 study The Use of Bigrams to Enhance Text Categorization [53], the Long Short-Term Memory (LSTM) model is particularly efficient when it comes to text classification and language translation tasks based on time series data [54]. Indeed, LSTM builds upon rudimentary Recurrent Neural Networks (RNNs) through an enhanced memory cell and three main gates - input, output

and forget - to keep the practical information and drop the impractical information. The concept of the Siamese network relies on architecture that consists of at least two identical sub-networks to lessen the model complexity by dividing the parameters. In Figure 2, the Siamese LSTM model outline is presented:

Where each pair of input sequences are denoted by $(X_1^a, \ldots, X_T^a)$ and $(X_1^b, \ldots, X_T^b)$, and the response labeled with y, at the first level all the input sequences have to be the same length where this can be done by padding. The result is then added to the LSTM layers.

3.2.2 BERT Architecture

A transformer-based architecture, Bidirectional Encoder Representations from Transformer (BERT) [55] is used to solve text classification and categorization tasks. BERT uses one-directional training to read input sequences from right to left or left to right in order to generate or predict the next word, Additionally, it also utilizes bidirectional or non-directional training [56] to produce a deeper sense of language context by taking both the previous and next tokens at the same time. The two major phases to BERT model: pre-training and fine-tuning. Basically, this means that the model is trained using unlabeled data from different pre-training tasks during the pre-training step. It implements all the pre-trained parameters and adjusts (or "fine tunes") them by utilizing labeled data gathered from downstream tasks during the fine-tuning step.

In this research, a BERT-base was adopted as the base model which consists of an encoder containing 12-layer transformer blocks. As shown is figure3, each block includes 768-dimensional hidden layers and 12-head self-attention layers. Moreover, a technique implemented in the BERT model called a masked language model which prepends each sentence with the special token CLS and appends the special separation token SEP at the end of each sentence. Additionally, it replaces some words randomly with a MASK token. The SEP token aids the model in comprehending where the end of one input is and where another one starts in the same sequence input. On the other hand, the special classification token CLS is included as the first token of every input sequence and is utilized in classification tasks as an aggregate of the entire sequence representation. The MASK token is used in the sentence in place of a desired predicted word.

In the end, using the following equation, a typical SoftMax layer is added to the top of the BERT model to forecast the label probability where *c* is the probability label, *s* is the whole sentence, *h* is the final hidden state of the first token, *b* is the bias vector and *w* is the weight matrix:

$$p(c|s) = softmax(W.h + b) \quad (1)$$

4. EXPERIMENTS

4.1 Dataset

To successfully achieve an accurate system with superior performance, Deep Learning models (including LSTMs and BERT) usually require large datasets due to their huge number of parameters. In this study, the researcher collected 206,080 tweets using Twitter API to create an in-house news dataset which was divided into two datasets as shown in Figure 4. The clean dataset contains 159,284 legitimate tweets and the fake dataset contains 46,796 fake tweets collected from the Anti-Rumor Authority which was founded in 2012. A sample of the collected tweets can be seen in Table 2:

4.2 Performance Measures

This study utilized the Confusion Matrix (True Positive, True Negative, False Positive, False Negative) to correctly classify instances to evaluate the proposed system's effectiveness and performance. These were then calculated based on the equation below [57]:

$$Accuracy = \frac{(TP + TN)}{TP + FN + FP + TN} \quad (2)$$

Where TP is True positive rate $TP_{rate} = \frac{TP}{TP+FN}$ the percentage of positive instances correctly classified.

TN is True negative rate $Tn_{rate} = \frac{TN}{FB+TN}$ is the percentage of negative instances correctly classified.

FP False positive rate $FP_{rate} = \frac{FP}{FB+TN}$ the percentage of negative instances misclassified.

FN False negative rate $FN_{rate} = \frac{FN}{TP+FN}$ the percentage of positive instances misclassified.

4.3 Experimental Setup

4.3.1 Shallow Learning

In shallow learning experiments, researchers randomly divided the training sets to 75% and testing sets to 25%. The shallow learning model was implemented using the python language, with the Hyper-Parameter tuning being implemented with regards to the training dataset and 5-fold cross validation.

4.3.2 Deep Learning

The train and test sets were also randomly divided into a split ratio of 75:25 in Deep learning experiments. In this study, researchers also utilized two different competitive deep learning models: LSTM and pre-trained BERT models which were carried out using PyTorch implementation along with the TensorFlow backend engine. For the BERT model, the batch size was initially set to 32, the learning rate to 2e-5, and epochs to 2. Additionally, for the LSTM model, pre-processing raw data was the input for the neural network model, the learning word embeddings with dimensionality d_in and LSTM sequence representation with dimensionality d_out being initialized between 10 and 50.

4.4 Results

Table 1 presents the results of this study on applying the shallow and deep classifiers on the dataset of the corpus in using the accuracy score metric to evaluate the classifiers.

Table 1: Experimental Results for Shallow and Deep Learning

CLASSIFIER	ACCURACY
SHALLOW LEARNING	95.92%
LSTM	96.71%
BERT	99%

A summary of the results is presented in both Table 3 and Figure 5 where it can be noticed that, when applying the pre-trained BERT model, this analysis attains a better testing accuracy than other initialization procedures. As shown above, an accuracy rating of 95.92% was obtained when shallow learning was applied to the pre-processed dataset. After applying the LSTM and BERT models to the same data, the accuracy improved to 96.71% and 99%, respectively. See figure 6.

4.5 Discussion

This sub-section compares information between two parameters - accuracy and loss - created by the deep and shallow method classifiers during the training and testing processes. Model error was calculated after each training session through the loss function. It then used the back propagation algorithm to update the threshold and weights to attain the most favorable deep learning model. In this study, 25 epochs were used to enable the LSTM and shallow learning model to learn the optimal parameters. In terms of lower error loss rates, researchers demonstrated the superiority of the deep learning approach through their diagrams.

in Figure8 where the training model can automatically present the feature importance estimation from a trained predictive model. This feature importance presents a score that signifies how useful each feature was in creating the classifier model. It is noticeably clear that the more often a feature is used to make key decisions, the higher its relative importance.

From the previous literature review on related work, comparisons and discussion of shallow learning models and deep learning models, we finally come to a conclusion that deep learning has outperformed the shallow learning models with regards to fake news detection based on Arabic language. The proposed model resulted in an accuracy rating of 99%.

5. CONCLUSION

The classification of fake news in a language with limited resources, such as Arabic, is critical for future research. As a result, in this study, a new classification system for detecting fake news in Arabic was proposed. Furthermore, a new in-house dataset has been created to evaluate both shallow and deep classifiers. The dataset contained 206,080 tweets obtained via the Twitter API. This dataset was used to ensure an accurate system with superior performance when shallow and deep learning models were employed. The study also focused on using shallow learning and comparing it to deep learning models in order to provide additional research on the performance of shallow learning on the Arabic language.

The results of applying the models to the dataset demonstrated that using a pre-trained BERT model resulted in an accuracy rating of 99%, outperforming all other classifiers.

The other models, shallow learning and LSTM, achieved accuracy of 95.92% and 96.71%, respectively. Future research focusing on sentiment analysis and applying ensemble techniques for detecting Arabic fake news, as well as collecting data from various social media platforms, is highly anticipated.

Acknowledgment:

The research reported in this publication was supported by the Deanship of Scientific Research and Innovation at Al-Balqa Applied University in Jordan (Grant Number: **DSR-2020#228**).

Funding:

The research reported in this publication was funded by the Deanship of Scientific Research and Innovation at Al-Balqa Applied University, Al-Salt, Jordan. (Grant Number: **DSR-2020#228**).

REFERENCES:

[1] Radcliffe, D. and P. Bruni, State of social media, Middle East: 2018. 2019.
[2] Domm, P., False rumor of explosion at White House causes stocks to briefly plunge; AP confirms its Twitter feed was hacked. In CNBC.COM. 2013.
[3] Jin, Z., et al. News credibility evaluation on microblog with a hierarchical propagation model. in 2014 IEEE International Conference on Data Mining. 2014. IEEE.
[4] Jin, Z., et al. Detection and analysis of 2016 us presidential election related rumors on twitter. in International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. 2017. Springer.
[5] Liaw, A. and M. Wiener, Classification and regression by randomForest. R news, 2002. **2**(3): p. 18-22.
[6] Alazab, M., et al., COVID-19 prediction and detection using deep learning. International Journal of Computer Information Systems and Industrial Management Applications, 2020. **12**: p. 168-181.
[7] Mahlous, A.R. and A. Al-Laith, Fake news detection in Arabic tweets during the COVID-19 pandemic. Int J Adv Comput Sci Appl, 2021.
[8] Thaher, T., et al., Intelligent Detection of False Information in Arabic Tweets Utilizing Hybrid Harris Hawks Based Feature Selection and Machine Learning Models. Symmetry 2021, 13, 556. 2021, s Note: MDPI stays neutral with regard to jurisdictional claims in published ….
[9] Al Zaatari, A., et al. Arabic corpora for credibility analysis. in Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). 2016.
[10] Alkhair, M., et al. An arabic corpus of fake news: Collection, analysis and classification. in International Conference on Arabic Language Processing. 2019. Springer.
[11] Sarah Alanazi , M.K., Arabic Fake News Detection In Social Media Using Readers' Comments: Text Mining Techniques In Action. IJCSNS International Journal of Computer Science and Network Security, 2020. **20**.
[12] Nagoudi, E.M.B., et al., Machine generation and detection of arabic manipulated and fake news. arXiv preprint arXiv:2011.03092, 2020.
[13] Al-Yahya, M., et al., Arabic Fake News Detection: Comparative Study of Neural Networks and Transformer-Based Approaches. Complexity, 2021. **2021**.
[14] Harrag, F., et al., Bert transformer model for detecting Arabic GPT2 auto-generated tweets. arXiv preprint arXiv:2101.09345, 2021.
[15] Saadany, H., E. Mohamed, and C. Orasan, Fake or real? A study of Arabic satirical fake news. arXiv preprint arXiv:2011.00452, 2020.
[16] Elhadad, M.K., K.F. Li, and F. Gebali. COVID-19-FAKES: A Twitter (Arabic/English) dataset for detecting misleading information on COVID-19. in International Conference on Intelligent Networking and Collaborative Systems. 2020. Springer.
[17] Haouari, F., et al., ArCOV19-rumors: Arabic COVID-19 twitter dataset for misinformation detection. arXiv preprint arXiv:2010.08768, 2020.
[18] Khouja, J., Stance prediction and claim verification: An Arabic perspective. arXiv preprint arXiv:2005.10410, 2020.

[19] Almerekhi, H. and T. Elsayed. Detecting automatically-generated arabic tweets. in AIRS. 2015. Springer.
[20] Reis, J.C., et al., Supervised learning for fake news detection. IEEE Intelligent Systems, 2019. **34**(2): p. 76-81.
[21] Ahmed, H., I. Traore, and S. Saad. Detection of online fake news using n-gram analysis and machine learning techniques. in International conference on intelligent, secure, and dependable systems in distributed and cloud environments. 2017. Springer.
[22] Ozbay, F.A. and B. Alatas, Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A: Statistical Mechanics and its Applications, 2020. **540**: p. 123174.
[23] Horne, B. and S. Adali. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. in Proceedings of the International AAAI Conference on Web and Social Media. 2017.
[24] Jain, A., et al. A smart system for fake news detection using machine learning. in 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT). 2019. IEEE.
[25] Veyseh, A.P.B., et al. Rumor detection in social networks via deep contextual modeling. in Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2019.
[26] Ghanem, B., et al., Fakeflow: Fake news detection by modeling the flow of affective information. arXiv preprint arXiv:2101.09810, 2021.
[27] Zhou, X., J. Wu, and R. Zafarani, [... formula...]: Similarity-Aware Multi-modal Fake News Detection. Advances in Knowledge Discovery and Data Mining, 2020. **12085**: p. 354.
[28] Wani, A., et al. Evaluating deep learning approaches for Covid19 fake news detection. in International Workshop on Combating On line Ho st ile Posts in Regional Languages dur ing Emerge ncy Si tuation. 2021. Springer.
[29] Nasir, J.A., O.S. Khan, and I. Varlamis, Fake news detection: A hybrid CNN-RNN based deep learning approach. International Journal of Information Management Data Insights, 2021. **1**(1): p. 100007.
[30] Rashkin, H., et al. 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. 2017.
[31] Castelo, S., et al. A topic-agnostic approach for identifying fake news pages. in Companion proceedings of the 2019 World Wide Web conference. 2019.
[32] Shu, K., et al., Fakenewsnet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media. arXiv preprint arXiv:1809.01286, 2018.
[33] Pérez-Rosas, V., et al., Automatic detection of fake news. arXiv preprint arXiv:1708.07104, 2017.
[34] Ghanem, B., P. Rosso, and F. Rangel, An emotional analysis of false information in social media and news articles. ACM Transactions on Internet Technology (TOIT), 2020. **20**(2): p. 1-18.
[35] Yang, Z., et al. Hierarchical attention networks for document classification. in Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. 2016.
[36] Beltagy, I., M.E. Peters, and A. Cohan, Longformer: The long-document transformer. arXiv preprint arXiv:2004.05150, 2020.
[37] Pennebaker, J.W., et al., The development and psychometric properties of LIWC2015. 2015.
[38] Simonyan, K. and A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[39] Jin, Z., et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. in Proceedings of the 25th ACM international conference on Multimedia. 2017.
[40] Alazab, M. and L.M. Batten, Survey in smartphone malware analysis techniques. New threats and countermeasures in digital crime and cyber terrorism, 2015: p. 105-130.
[41] Alazab, M., A. Alazab, and L. Batten. Smartphone malware based on synchronisation vulnerabilities. in ICITA

2011: Proceedings of the 7th International Conference on Information Technology and Applications. 2011. ICITA.
[42] Batten, L.M., V. Moonsamy, and M. Alazab, Smartphone applications, malware and data theft, in Computational intelligence, cyber security and computational models. 2016, Springer. p. 15-24.
[43] Alazab, M., et al. Analysis of malicious and benign android applications. in 2012 32nd International Conference on Distributed Computing Systems Workshops. 2012. IEEE.
[44] Moonsamy, V., M. Alazab, and L. Batten, Towards an understanding of the impact of advertising on data leaks. International journal of security and networks, 2012. **7**(3): p. 181-193.
[45] Alazab, M., Automated malware detection in mobile app stores based on robust feature generation. Electronics, 2020. **9**(3): p. 435.
[46] Alazab, A., et al. Using feature selection for intrusion detection system. in 2012 international symposium on communications and information technologies (ISCIT). 2012. IEEE.
[47] Alazab, M., et al., Cybercrime: the case of obfuscated malware, in Global security, safety and sustainability & e-Democracy. 2011, Springer. p. 204-211.
[48] Alazab, M., et al., Information security governance: the art of detecting hidden malware, in IT security governance innovations: theory and research. 2013, IGI Global. p. 293-315.
[49] Alazab, A., et al. Web application protection against SQL injection attack. in Proceedings of the 7th International Conference on Information Technology and Applications. 2011.
[50] Alazab, M.V., Sitalakshmi ; Watters, Paul ; Alazab, Moutaz. , Zero-day malware detection based on supervised learning algorithms of API call signatures., in 9th Australasian Data Mining Conference. 2010.
[51] Moh'd A Mesleh, A., Chi square feature extraction based svms arabic language text categorization system. Journal of Computer Science, 2007. **3**(6): p. 430-435.
[52] Krizhevsky, A., I. Sutskever, and G.E. Hinton, Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012. **25**: p. 1097-1105.
[53] Tan, C.-M., Y.-F. Wang, and C.-D. Lee, The use of bigrams to enhance text categorization. Information processing & management, 2002. **38**(4): p. 529-546.
[54] Hochreiter, S. and J. Schmidhuber, Long short-term memory. Neural computation, 1997. **9**(8): p. 1735-1780.
[55] Devlin, J., et al., Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
[56] Vaswani, A., et al. Attention is all you need. in Advances in neural information processing systems. 2017.
[57] Alazab, M., et al., Intelligent mobile malware detection using permission requests and API calls. Future Generation Computer Systems, 2020. **107**: p. 509-521.

Table 2 comparison table between the Fake News Detection Models Using ML & DL literatures review

Reference	Language	Learning Type	Dataset	Method	Classifierz	Accuracy
[7]	Arabic	Shallow Learning	Collected 5.5 million Arabic tweets	Proposed a model that addressed the issue of classifying fake news about the covid-19 in Arabic,	NB, LR, SVM, MLP, RF, and XGB	93.3%
[8]	Arabic	Shallow Learning	Syrian crisis tweets [9]	Proposed an intelligent detection system for Arabic fake Twitter corpus news using eight ML classifiers and HHO feature selection approach	NB, KNN, SVM, DT, LR, LDA, RF, and XGBoost	81.5%
[10]	Arabic	Shallow Learning	Collected 4079 from YouTube comments	Investigated the fake news content in the Middle East through the Arabic comments on YouTube	SVM, DT, MNB	95.35%
[11]	Arabic	Shallow Learning	Build their own dataset	proposed a system for Arabic Fake news classification by analyzing the peoples comments on social media news using text mining techniques	DT, SVM, NB, KNN	87.18%
[12]	Arabic	Deep Learning	AraNews, ATB, Khouja	Developed models to detect the manipulated Arabic news that potentially fake	mBERT, AraBERT, XLM-RBase, XLM-RLarge	89.23%
[13]	Arabic	Deep Learning	Covid-19-Fakes, AraNews, ArCOV19-Rumors and ANS corpus	Presented a comparative study of neural network and transformer-based language models by examined them for Arabic fake news detection	Deep learning models (CNN, RNN, GRU) and Transformer-based models (Marbert, Arbert, ArElectra, QARiB, (AraBERT v1, v02, v2)	95%
[14]	Arabic	Deep Learning	Collecting 4196 tweets	proposed a transfer learning-based model to detect the Arabic sentences that were wrote by bots to check if it's fake or not	LSTM, BI-LSTM, GRU, BI-GRU and AraBERT	98.7%
[15]	Arabic	Deep Learning	Collecting 3185 fake news and 3710 real news	Combined the CNN model with pretrained word embeddings to proposed a fake news detection model	NB, XGBoost, CNN	98.59%
[20]	English	Shallow Learning	2282 BuzzFeed news articles	presented a new set of features for training classifiers and measured the prediction of current approaches and features performance for automatic fake news detection	KNN, NB, RF, SVM, XGB	86%
[21]	English	Shallow Learning	Collected 12,600 truthful articles and 12,600 fake articles, in addition to Horne & Adali dataset	presented an online fake news classification model that used n-gram method and machine learning	SVM, LSVM, KNN, DT, SGD, LR	92%
[22]	English	Shallow Learning	BuzzFeed Political News Data set, Random Political News Data set, ISOT Fake News data set	proposed a two-steps method to identify fake news on social media	23 different supervised classifiers	96.8%

[24]	English	Shallow Learning	Build their own dataset	demonstrated a new methodology for fake news detection that consist of: Aggregator, News Authenticator, News Suggestion	SVM, NB	93.6%
[25]	English	Deep Learning	Twitter 15 and Twitter 16	Proposed a semantic graph model to detect the rumors that spread in Twitter.	Semantic Graph	76.8%
[26]	English	Deep Learning	TruthShades, PoliticalNews, FakeNewsNet, and MultiSourceFake	proposed the FakeFlow system to detect the fake news articles by implementing the flow of affective data in a neural architecture.	FakeFlow model	96%
[27]	English	Deep Learning	PolitiFact, GossipCop	proposed SAFE model to investigated the role of the similarity between news visual and textual content for fake news detection	SAFE model	87.4%
[28]	English	Deep Learning	Contraint@AAAI 2021 Covid-19	Evaluated different supervised text classification algorithm	CNN, LSTM, BERT	98.41%
[29]	English	Deep Learning	ISOT and FA-KES	Presented a hybrid deep learning model that contains a combination of complex and recurrent neural networks to solve the fake news classification problem	CNN and LSTM	99%
Our Proposed model	Arabic	Shallow Learning	Collected 206080 Arabic tweets	Proposed a model that addressed the problem of detecting the Arabic fake news	LSTM and BERT	99%

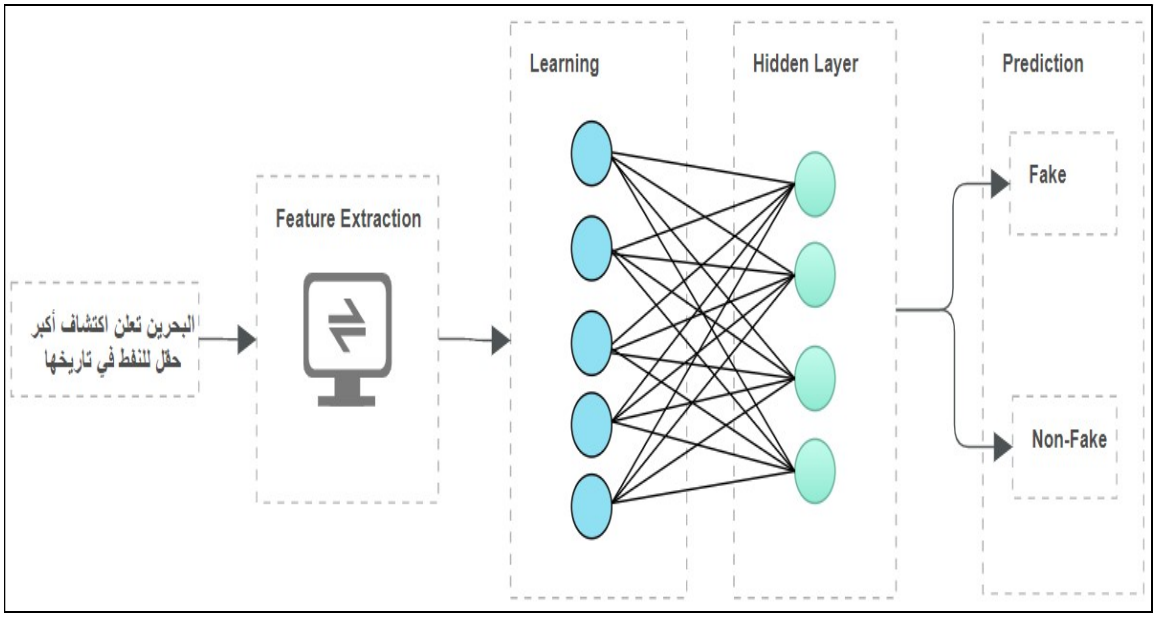

Figure 1: The Proposed Shallow Learning Architecture

Journal of Theoretical and Applied Information Technology
15th March 2023. Vol.101. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

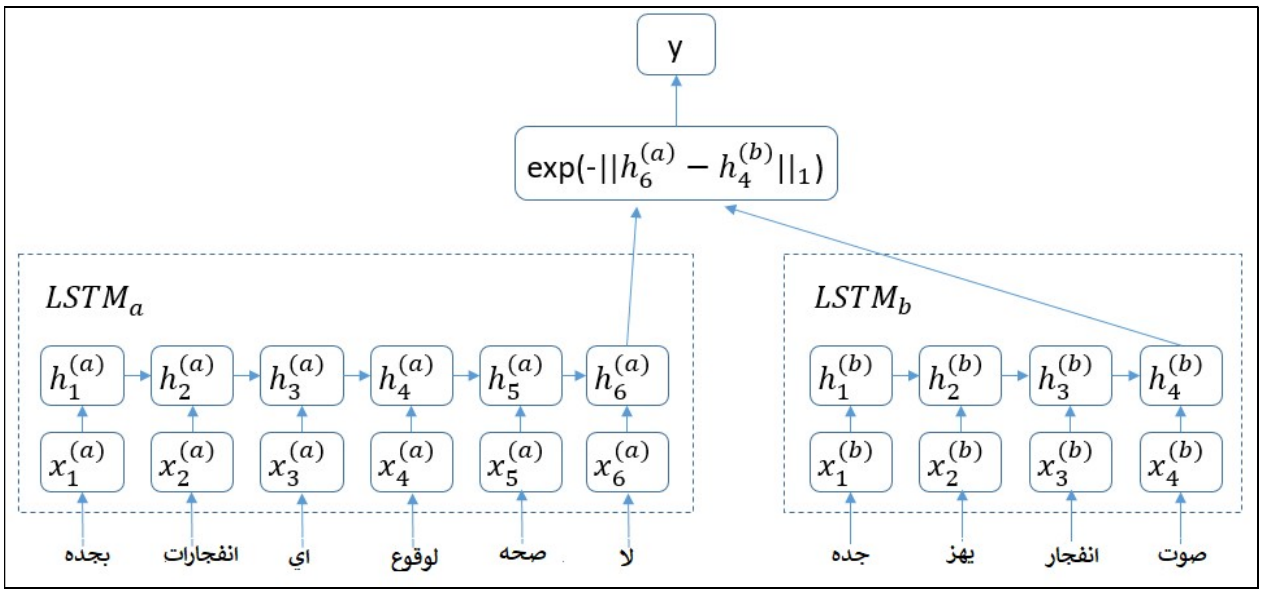

Figure 2: The Proposed Siamese LSTM model

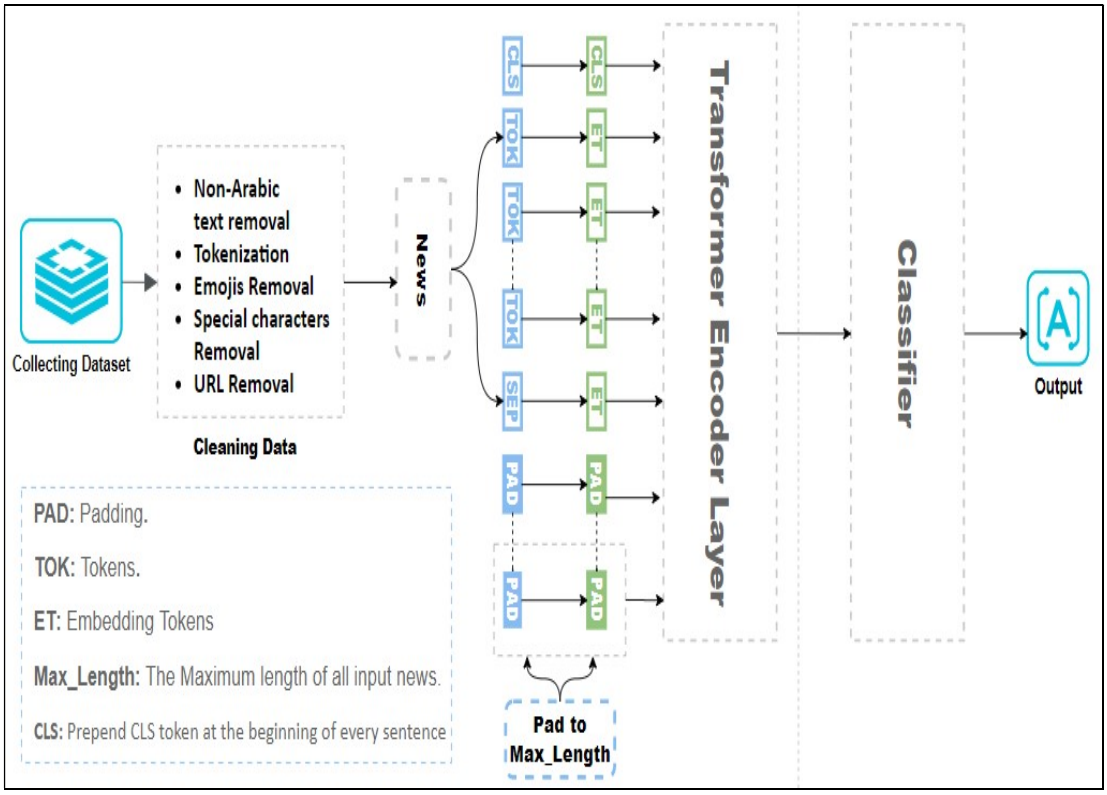

Figure 3: The Proposed BERT model

Journal of Theoretical and Applied Information Technology
15th March 2023. Vol.101. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3: A Sample of the Collected Dataset

News Type	News
Fake	انتقل الى رحمة الله تعالى الممثل ناصر القصبي بعد تعرضه الى حادث على طريق القصيم الله يرحمه وفاة ناصر القصبي
Fake	عاجل | السعوديه بسبب سوء الاحوال الجويه هبوط طائرة قبل قليل بين مكة و جدة على الخط السريع
Not Fake	الخطوط السعودية تعلن تسيير رحلات يومية مباشرة إلى موسكو تزامنا مع مونديال كأس العالم 2018م في روسيا
Not Fake	دراسة: الحديد في الجسم يزيد نسبة التحصيل العلمي للأطفال

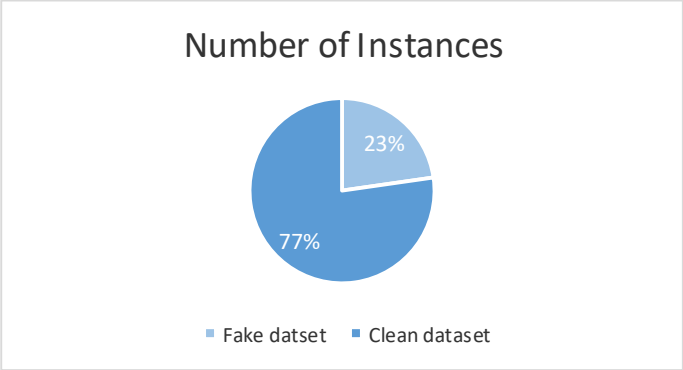

Figure 4: Statistics of Dataset

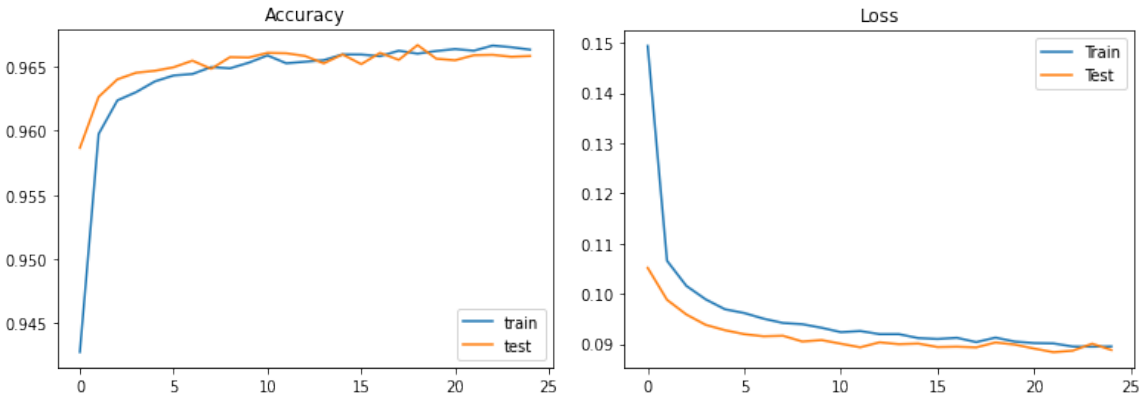

Figure 5: The Accuracy and Loss Charts for Proposed Deep Learning Model

Journal of Theoretical and Applied Information Technology
15th March 2023. Vol.101. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

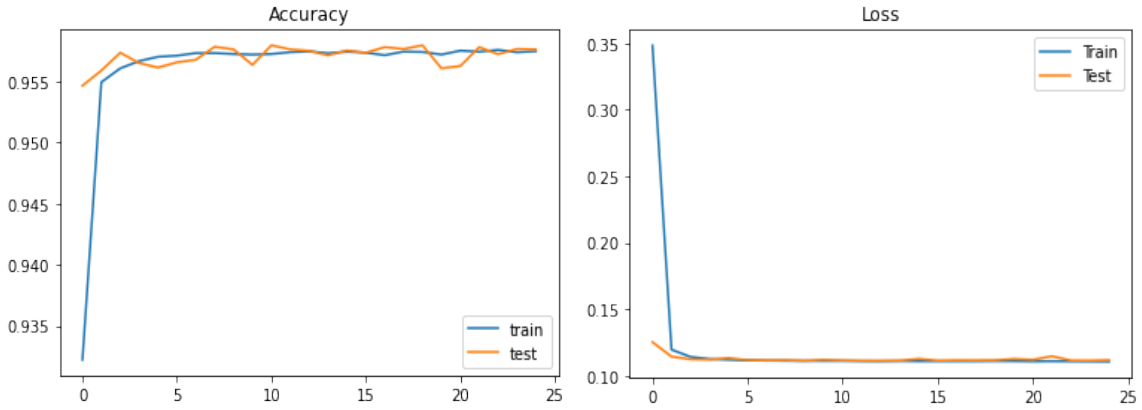

Figure 6: The Accuracy and Loss charts for Proposed Shallow Learning model

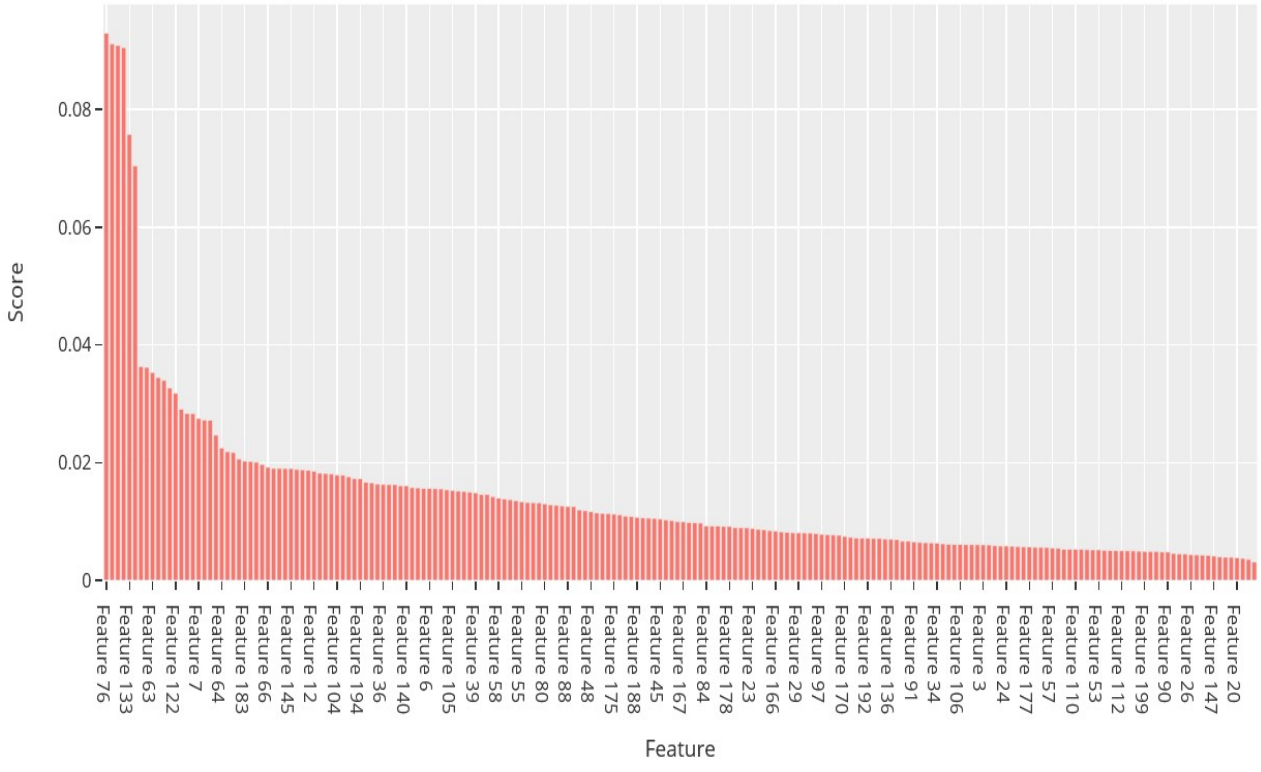

Figure 7: Feature Importance for the Proposed Shallow Learning Model