

Automatic Fake News Detection based on Deep Learning, FastText and News Title

Youssef Taher¹

Center of Guidance and Planning
Rabat, Morocco

Adelmoutalib Moussaoui²

Faculty of Sciences and Technologies
Moulay Ismail University
Errachidia, Morocco

Fouad Moussaoui³

FST of Errachidia, Moulay Ismail
University, Errachidia, Morocco

Abstract—As a range of daily phenomena, Fake News is quickly becoming a longstanding issue affecting individuals, public and private sectors. This major challenge of the connected and modern world can cause many severe and real damages such as manipulating public opinion, damaging reputations, contributing to the loss in stock market value and representing many risks to the global health. With the fast spreading of online misinformation, checking manually Fake News becomes ineffective solution (not obvious, difficult and takes a long time). The improvement of Deep Learning Networks (DLN) can support with high degree of accuracy and efficiency the classical processes of Fake News spotting. One of the keys improvement strategies are optimizing the Word Embedding Layer (WEL) and finding relevant Fake News predicting features. In this context, and based on six DLN architectures, FastText process as WEL and Inverted Pyramid as News Articles Pattern (IPP), the present paper focuses on the assessment of the first news article feature that is hypothesized as affecting the performances of fake news predicting: News Title. By assessing the impact that the Embedding Vector Size (EVS), Window Size (WS) and Minimum Frequency of Words (MFW) in News Titles corpus can have on DLN, the experiments carried out in this paper showed that the News Title feature and FastText process can have a significant improvement on DLN fake news detection with accuracy rates exceeding 98%.

Keywords—Fake news; automatic detection; deep learning; FastText; news title

I. INTRODUCTION

Nowadays, fake news and sophisticated disinformation can have serious real world negative effects [1]. Often the main objectives of these false information is to intentionally deceive, gain attention, manipulate public opinion or to damage reputations.

During a time of uncertainty or crisis (ex. Covid-19), people are more likely to believe the false rumour they find on web, social media as well as online newspapers if it appeals to their emotions. As shown in Table I, this phenomenon of fake news can fall into different categories and take on different faces.

Deciphering these massive, instantaneous and heterogeneous daily news categories are valid or not becomes a serious challenge. In the era of social media, these false information and hoaxes spread freely, wider and more faster than ever before. These digital platforms enable novel forms

of communication, affect and accelerate the way individuals interpret daily developments.

Recently, in many democratic systems fake news distort and change how the electoral campaigns of candidates and political parties [2]. In the post-election period, significant number of websites and social media publish and share falsified or heavily biased information and stories, which calls into question the legitimacy of the elections.

In business and economics systems fake news and sophisticated disinformation are currently a hot topics. On the world economy, disinformation and hoaxes have direct and greater impacts. As example of consequences, every year fake news contribute to the loss in stock market value and investors can lose money due this problem [3].

In health systems, false information can spread easily and widely than pandemics, and can become accepted as true [4]. False information about virus can do real harm with often dangerous consequences. As examples, misleading information about medical services, medical products, treatments, official sources and guidelines can directly endangering the public health.

To develop the ability to decipher these fake news categories, various techniques and approaches have been developed. These solutions can be classified into two classes of methods (Fig. 1).

The first class is manual. This class is mainly based on comparing real news with unverified news by visiting fact checking sites.

TABLE I. FAKE NEWS CATEGORIES

Fake news categories	
▪	Blatantly false articles
▪	True articles with some false interpretations
▪	Pseudoscientific articles
▪	Opinion articles disguised as news
▪	Satirical articles
▪	Articles comprised of mostly text, tweets and quotes from other people

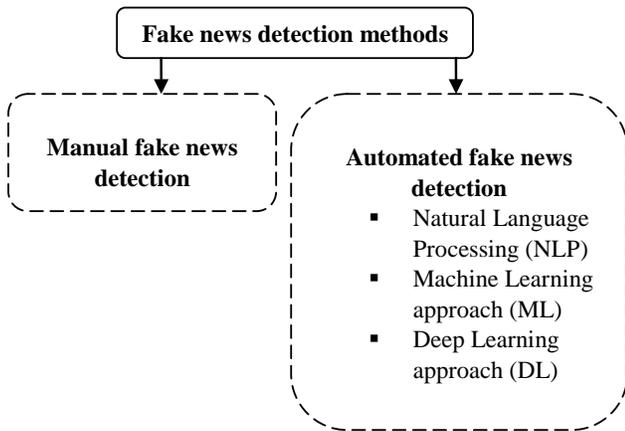


Fig. 1. Fake News Detection Approaches.

The second class is based on automated systems. To detect and predict misinformation, these recent systems exploit many important benefits of Natural Language Processing (NLP) and Artificial Intelligence (AI) [5] such as Machine Learning (ML) algorithms and Deep Learning (DL) networks (Singh et al., 2021)[6].

Recently, the improvement of these automated models can support with high degree of accuracy and efficiency the classical process of Fake News spotting. One of the key improvement strategies is optimizing the main Fake News predicting features and Word Embedding Layer (WEL).

In this context, and based on many DLN architectures and the news articles pattern IPP (Inverted Pyramid Pattern), the present paper focuses on the assessment of the first news feature that is hypothesized as affecting the performances of automatic Fake News spotting: News articles title.

The experiments carried out in this paper assess the impact of this key feature on Fake News DLN predicting performances by using six Deep Learning models: Simple LSTM, Stacked LSTM with two layers, Bidirectional LSTM, Simple GRU, Stacked GRU with two layers and Bidirectional GRU. These DLN models are feed by several rows of real and Fake News titles. Each title is a collection of English language words. To improve the embedding process of this corpus, the FastText process is used as a first embedding layer.

For each used model, the experiments assess the impact that Embedding Vector Size (EVS), Window Size (WS) and Minimum Frequency of Words (MFW) in the used corpus can have on Fake News spotting performances (execution time, loss and accuracy). To illustrate the diagnostic ability of each used model as its discrimination threshold is varied, the Receiver Operating Characteristics Curve (ROC) are plotted and the AUC values (Area Under the ROC Curve) are calculated by testing several values of EVS, WS and MFW.

Compared to many recent relevant studies on automatic fake detection based on the DLN architectures (summarized in paragraph III-B), the main objective of these experiments is to improve the performance of the execution time, loss and accuracy by testing a new embedding process (FastText), and reducing the dimensionality of the used articles news data by assessing the impact of the first news title feature that is

hypothesized as affecting the performances of automatic Fake News spotting.

This paper is organized as follows. Section 2 presents a brief overview of the used deep learning architectures. Section 3 provides a review of recent relevant studies on Fake Detection based on DLN and we summarize our findings. The contribution of this paper is presented in Section 4. This section presents the used architecture, embedding layer, used dataset, summarizes the five main features of the Inverted Pyramid Pattern (IPP), and discuss the impact of news title feature and FastText on fake news spotting performances. Finally, Section 5 presents our conclusions.

II. FUNDAMENTALS OF THE USED DEEP LEARNING ARCHITECTURES

In Deep Learning, Recurrent Neural Networks (RNN) are a family of Neural Networks [7]. These networks excels in learning by processing sequential data (one input follows another in time). RNN models use the current input and remember the preceding elements. The output at the current time step becomes the input to the next time step (Fig. 2).

Through its effectiveness, this kind of neural network is often used to handle text as news articles, tweets, comments, and have shown an important success in many Natural Language Processing (NLP) projects (Machine Translation, Speech Recognition, Generating Image Descriptions).

The basic architecture of RNN networks is Vanilla RNN (RNN network with single hidden layer). To deal with many limits of this RNN class, researchers have invented more advanced types of RNNs [8] such as Stacked RNNs, Bidirectional RNNs, Deep Bidirectional RNNs, Long Short Term Memory Networks (LSTM) [9], and Gated Recurrent Unit Networks (GRU).

The Stacked (Deep) RNN networks use multiple hidden RNN layers (Fig. 3). This architecture stacks multiple layers on top of each other. Each layer contains multiple cells, processes some part of the project tasks and passes it on to the next layer. The last layer provides the output. This approach (processing pipeline) has many potential benefits: exponentially more efficient to represent some functions and can for example extract more abstract features of news titles. However, these networks suffer from the vanishing gradient problem in the vertical direction.

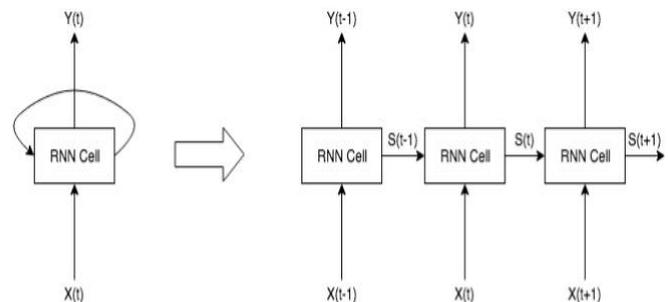


Fig. 2. Vanilla RNN.

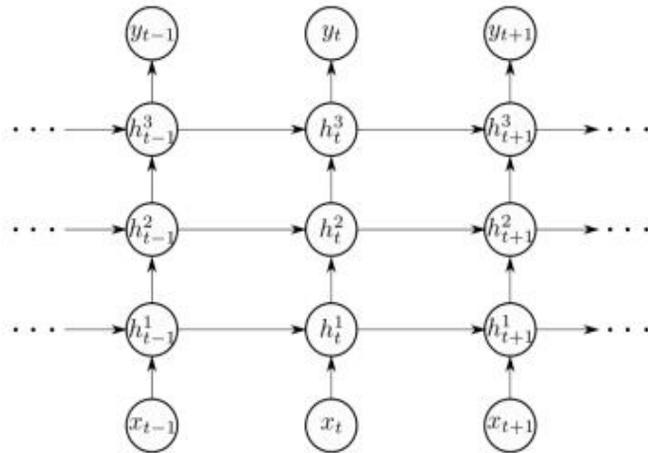


Fig. 3. Stacked RNN.

Bidirectional Recurrent Neural Networks (BRNNs) (Fig. 4) put two independent RNNs together without interacting with one another [10]. The first RNN network feeds the input sequence in normal time order (positive time direction). The second one feeds the input in reverse time order (negative time direction). Therefore, the model receives information from both past and future states. At each time, the output of BRNNs is computed after passing the merged results (by concatenating, adding, multiplying) of the forward and backward layers into the sigmoid function. Applied to fake news spotting process, this approach can improve the model performance and accuracy by obtaining the context in two directions compared to unidirectional RNN.

The Recurrent Neural Networks presented above suffer from vanishing and exploding gradient problems [11]. One of the important approaches to deal with these problems is Long Short Term Memory Networks (LSTM) [12]. To learn, remember and store relevant information for learning, these networks use a memory unit. By using a gating mechanism, LSTM architecture decides to pass the information to the next layer or forget the information it has. Fig. 5 summarizes the basic cell of LSTM network.

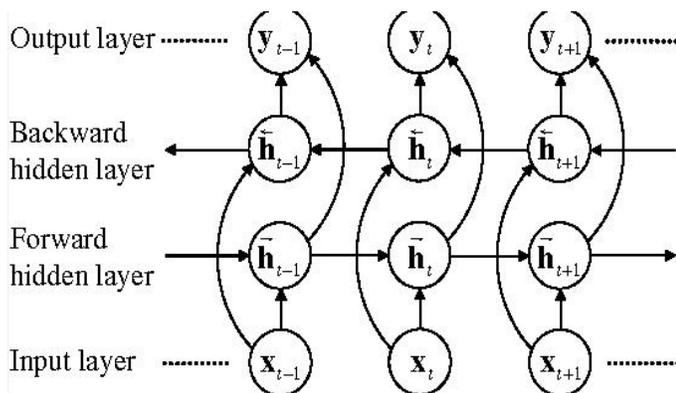


Fig. 4. Bidirectional RNN.

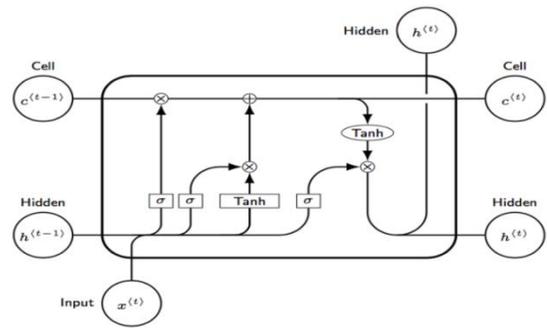


Fig. 5. Basic Cell of LSTM Network.

Each LSTM cell has three inputs ($h(t-1)$, $c(t-1)$, $x(t)$) and two outputs ($h(t)$, $c(t)$).

$x(t)$: the input at time t .

$h(t-1)$: the previous hidden state.

$h(t)$: the current hidden state.

$c(t-1)$: the previous cell state (the explicit memory unit).

$c(t)$: the current cell state.

The forget gate (first sigmoid layer with the inputs $x(t)$ and $h(t-1)$) selects the amount of information of the previous cell to be included. The input gate (second sigmoid layer with the inputs $x(t)$ and $h(t-1)$) decides what new information is to be added to the cell. These sigmoid layers determine the information to be stored in the cell state. By calculating the point-wise multiplication of the result of the tanh layer and the result of the input gate (Fig. 5), LSTM cell decides the amount of information to be added to the cell state. To produce $c(t)$, the result of this point-wise multiplication is added with the result of the first sigmoid layer multiplied by $c(t-1)$. By using a sigmoid and a tanh layer, the LSTM cell calculates the output. The sigmoid layer decides which part of $c(t)$ will be present in the output. The tanh layer shifts the output in the range of $[-1,1]$.

To simplify the internal design and to improve the design complexity, Cho proposed the Gated Recurrent Unit Network (GRU). This variant of the LSTM is based on two gates illustrated in Fig. 6: update gate (z) and a reset gate (r). To keep around how much previous memory, the GRU cell uses the update gate. To define how much information needs to be forgotten this cell uses the reset gate.

z : update gate

r : reset gate

Instead of the input, forget, and output gates in LSTM cell, GRU architecture uses this gated mechanism to capture dependencies of different time scales effectively, and retains LSTM's resistance to the vanishing gradient problem. The internal structure of GRU needs few computations to make updates to its hidden state.

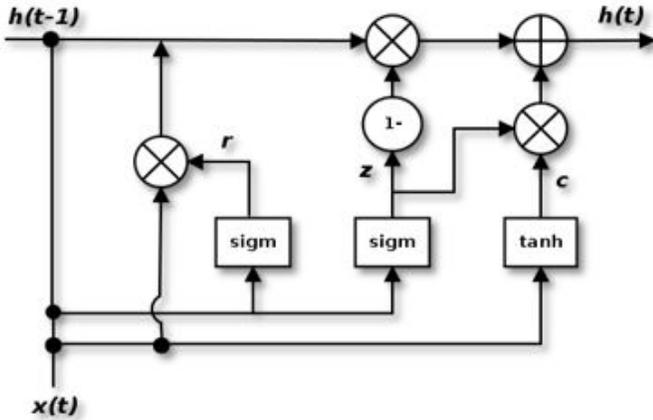


Fig. 6. Basic Cell of GRU Network.

III. RELATED WORK

A. Study Objective

This study investigates the impact of news articles titles on Fake News DLN spotting performances by using six DLN models: Simple LSTM (SI_LSTM), Stacked LSTM (ST_LSTM), Bidirectional LSTM (BI_LSTM), Simple GRU (SI_GRU), Stacked GRU (ST_GRU) and Bidirectional GRU (BI_GRU). To improve vectors representation of news titles corpus, the FastText library is used as first embedding layer.

By feeding these six models by several values of Embedding Vector Size (EVS), Window Size (WS) and Minimum Frequency of a Word in news titles corpus (MFW), an empirical study is performed on how does news titles impact Fake News prediction performances based on DLN.

To check and visualize these performances, the assessment process is based on the time execution, loss and accuracy. To illustrate the diagnostic ability of each used model as its discrimination threshold is varied, the Receiver Operating Characteristics curves (ROC) are plotted [13] and the Area under the ROC Curve values (AUC) are calculated [14].

The ROC curves are plotted with the rate TFNR (True Fake News Rate) on the y-axis against the rate FFNR (False Fake News Rate) on the x-axis. These rates are defined as follows:

$$TFNR = \frac{\text{True Fake News}}{\text{True Fake News} + \text{False Real News}} \quad (1)$$

$$FNR = \frac{\text{False Fake News}}{\text{True Real News} + \text{False Fake News}} \quad (2)$$

B. Automatic Fake News Detection based on Deep Neural Networks (DLN)

In the paragraph below, a summary of recent relevant studies on automatic fake detection based on the DLN architectures is provided above.

Recently, detecting fake news and sophisticated disinformation has become an important need and challenge for citizens and governments. This phenomena is turbo-charged by digital technology, and can have significant negative effects on individuals, social, political and economic environments.

Across the world, people need to be well-equipped to separate false information from real information. Using classical solutions (manual processes) to meet this need has many drawbacks. Indeed, with the fast spreading of online information, checking manually Fake News becomes an ineffective approach (not obvious, difficult, takes a long time). Therefore, automated Fake News detection based on Natural Language Processing (NLP), Machine Learning (ML) and Deep Learning (DL) present an efficient, accurate and fast solution to support and improve manual methods [15].

In the era of artificial intelligence, Deep Learning models can be trained through the use of large amounts of real / fake articles news and accomplishing complex news tasks. One of these important tasks is Fake News prediction.

Lastly, Deep Learning architectures such as Recurrent Neural Networks (RNN), Short Term Memory Networks (LSTM) and Gated Recurrent Unit Network (GRU) offer a lot of promise for spotting Fake News. Indeed, various recent research projects have been used these architectures to detect Fake News by taking advantages of these networks.

Among these recent investigations, we can cite the important recent study of S. R. Sahoo and B. B. Gupta [16]. In this investigation, the authors introduce automatic Fake News detection approach in chrome environment on which it can detect Fake News on Facebook. They use multiple features associated with Facebook account with some news content features to analyze the behavior of the account through deep learning. This Fake News detection approach has achieved higher accuracy than the existing state of art techniques.

Other recent important deep learning model is proposed by R. K. Kaliyar et al. [17]. To extract several features at each layer, the authors propose a Deep Convolutional Neural Network (FNDNet) for Fake News detection. The proposed model achieved state-of-the-art results with an accuracy of 98.36% on the test data.

To address the shortcoming caused by Deep Learning model entirely based on Natural Language Processing (NLP), D. S and B. Chitturi [18] propose a new Deep Neural approach to Fake News identification. This system includes a live data stage mining which provides secondary features (source domains of the article, author names, etc.). By exploring LSTM and FF Neural Networks, the authors seek to compare the results from models with and without these secondary mined features.

By using different embedding models for news items of different lengths, M. H. Goldani et al [19] propose the use of capsule neural networks in the fake news detection task. The authors use two recent well-known datasets in the field, namely ISOT and LIAR. They apply different levels of n-grams for feature extraction. The results show encouraging performance.

Based on Bi-directional LSTM-recurrent neural network, Bahad et al. [10] propose a deep leaning model for a fake news detection. This study uses two publicly available unstructured news articles datasets are used to assess the performance of the model. This model shows an important

accuracy over other methods namely CNN, vanilla RNN and unidirectional LSTM.

C. Used Embedding Process

Producing efficiently numerical dense vector (word embedding) is a key process for many news articles processing tasks such as articles news classification, fake news predicting, etc.

The optimized numerical vectors can encode efficiently the semantic information, measure the semantic similarity between two words in news articles, and use these numerical vectors as news articles features [20].

One of the most popular techniques used to create and to learn these vectors is Word2Vec [21]. This model developed by Google supports supervised learning and unsupervised learning. It based on two methods involving simple Neural Networks with one hidden layer: the Skip-Gram model and the Continuous Bag-of-Words model (CBOW). The word vectors are learned via backpropagation and stochastic gradient descent. The Skip-Gram model can use the target word to predict the context. The CBOW method takes the context of each word as the input and tries to predict the word corresponding to the context.

As a modified version (extension) of Word2Vec (Skip-Gram and CBOW) presented above, the FastText process [22] is used as embedding layer in the proposed architecture.

FastText is a library for efficient word embeddings and text classification. This library is developed by the Facebook research team has shown excellent results on many Natural Language Processing (NLP) projects (faster with superior performance).

To improve the used vector representations, FastText process treats each word in news titles corpus as composed of character n-grams (split each word in multiple n parts). It may be bigram, trigram, etc. (Table II). The character n-grams of length n can be generated by sliding a window of n-characters from the start till the end.

By providing Skip-Gram and CBOW models, FastText process computes News words representations. Each word of the used dictionary is represented by the sum of the vector representations of its n-gram. Consequently, and by averaging the vectorized representation of all its constituent n-grams, the embedding process can generate News Titles word vectors for the words that does not appear in the training corpus.

In addition to this last benefit, this used embedding layer is significantly better than the original Word2Vec on syntactic tasks, especially in the case of training corpus with small sizes (case of the present study).

The pooling strategy of FastText can generate a huge number N of unique n-grams. To bind the memory requirements, a hashing process is used. Each character n-gram is hashed to an integer between 1 to H (bucket size). Therefore, FastText learns a total H embeddings instead of learning a total N embedding (Fig. 7).

TABLE II. EXAMPLES OF DIFFERENT LENGTH CHARACTER N-GRAMS

Word	Length(n)	Character n-grams
author	3	<au, aut, uth, tho, hor, or>
author	4	<aut, auth, utho, thor, hor>
author	5	<auth, autho, uthor, thor>
author	6	<autho, author, uthor>

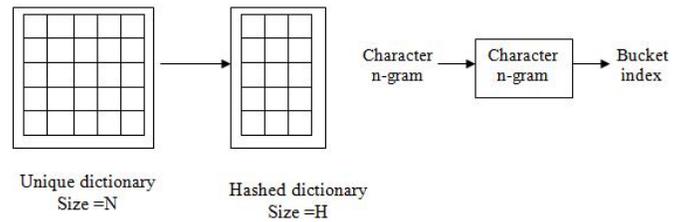


Fig. 7. Hashed Dictionary.

D. Used News Article Pattern

Usually, news articles describe events, persons, occurrences, experiences, places and other topics by following a particular pattern (how information should be prioritised and structured) [23].

One of the most common used patterns is the Inverted Pyramid (IP) which often composed of five important news features (Fig. 8).

News information in the IP pattern is presented in descending order of importance. The first part (feature) is the title (headline). This feature tells what the news is about. The second feature shows who wrote the news (byline). The third feature is the lead (first paragraph). This paragraph summarizes the main and important facts of the news, and based on 5 W's (who, what, when, where, and why) and how. The fourth feature is the body. This part is the core information and details about the news, which supports and amplifies the lead. The fifth feature is the ending which usually gives something to think about.

Often the order of IP pattern allows reading quickly the most crucial information, and estimating an initial manual spotting of Real or Fake News.

Based on IP order of importance, this paper assesses the impact of news title feature (first feature) that is hypothesized as affecting the performances of fake news spotting.

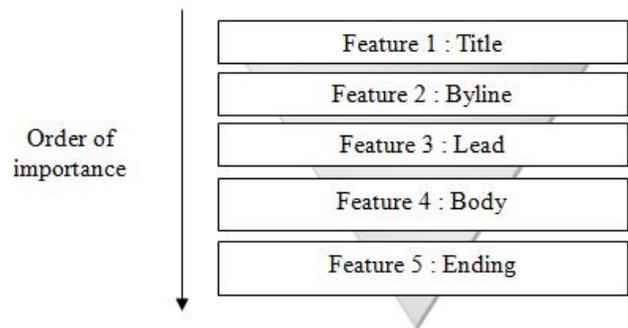


Fig. 8. Used Article News Pattern IP.

E. Used Dataset

The present study uses two Fake and Real News datasets from Kaggle source [24]. This news dataset is based on two 2 files (two lists of news articles): Fake News and True News files. Each file has four features: News Title, News Lead, News Subject and News Date (Table III).

TABLE III. FEATURES OF THE USED DATASET

News title	News lead	News subject	News date
The title of the article	The lead of the article	The subject of the article	The date at which the article was posted

After downloading, these two files are merged into a single dataset. This news dataset is labelled by adding a new feature called News Label. "0" value is assigned to Fake News (FN) and "1" value is assigned to True News (TN).

Fig. 9 summarizes the number of each label value. News Title data are extracted to fed FastText embedding layer and building the used six DLN models.

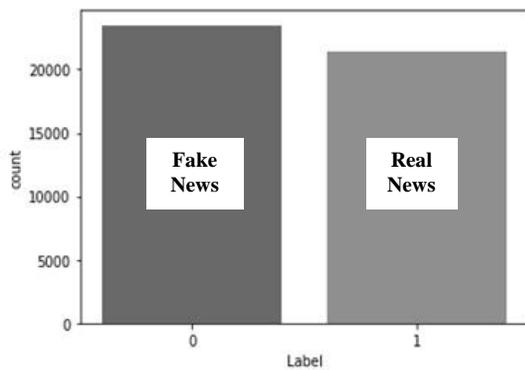


Fig. 9. Number of used Fake and True News Articles.

The encoded news data are tokenized, created and padded by using TensorFlow [25] and Keras [26] preprocessing tools (Fig. 10).

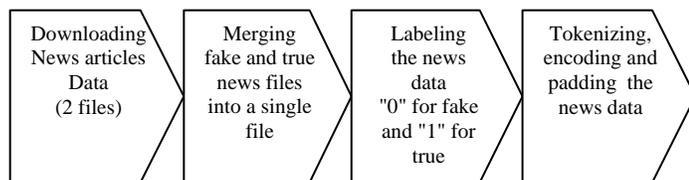


Fig. 10. Articles News Data Pre-Processing Process.

F. Automatic Fake News Detection based on DLN, FastText and News Titles

1) Assessment of execution time performance: Usually fake news predicting projects focus on the accuracy of Deep Learning (DL) or Machine Learning (ML) models that they are using. However, optimizing DLN execution time is one of the important processes that can improve the performances of Fake News detection software. Particularly, when DL model is used as a specific service or part of service installed on personal computers or other devices with limited resources [27]. In this context, the first experiments of this study assess

the time performances of the used embedding layer (FastText process) based on news title feature.

Generally, the duration of the embedding process can be due to the hardware platform architecture (Central Processing Unit (CPU), Graphics Processing Unit (GPU), Random Access Memory (RAM)), internal or external interruptions during the computation and the used libraries if are optimized or not.

The first experiments used the hardware configuration summarized in the following Table IV:

TABLE IV. USED HARDWARE CONFIGURATION

Processor	Intel (R) Core (TM) i7- 4610 M
Processor's speed	CPU@3.00 GHz
Memory size	8.00 Go

The main objective of these first experiments is estimating and assessing the impact that the main parameters of the embedding layer (Embedding Vector size (EVS), Window Size (WS), Minimum Frequency of Words (MFW)) can have on time performance. These experiments start by feeding the embedding layer by several EVS values and setting the used maximum dimension to 140 (≈ half of the maximum length of News Titles). As shown in Fig. 11, the execution time of the embedding layer increases with important rate by increasing EVS values. The average execution time obtained by smaller sizes (less than 40) is relatively small compared to longer embedding vectors (greater than 100). If the embedding size go from 10 to 100, the execution time will double its value.

The second experiment assesses the effect of WS parameter (window of surrounding context words) on the execution time. We fed the embedding layer by several WS sizes on a scale of 1 to 10. As shown in Fig. 12, as WS values increase, the execution time increases. When WS changes from WS =1 to WS = 10, the approximate change in the execution time is around 67 percent (67, 01%). Large values of WS (greater than WS = 5) increase slightly the execution time (≈3 seconds of difference from WS = 6 to WS = 10).

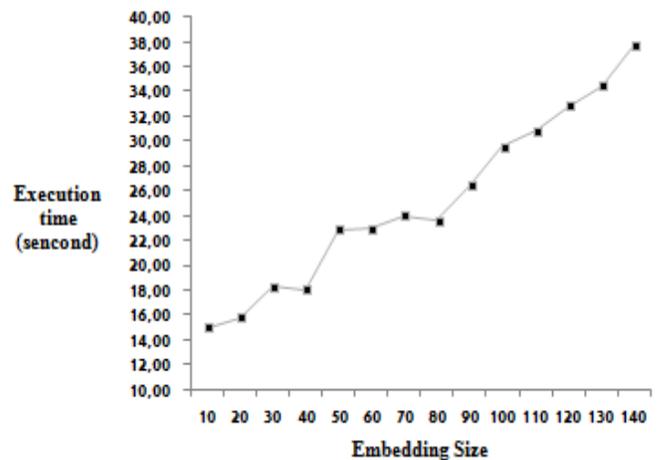


Fig. 11. Variation of the Execution Time Values with Different Embedding Vector Sizes. WS =2, MFW=2.

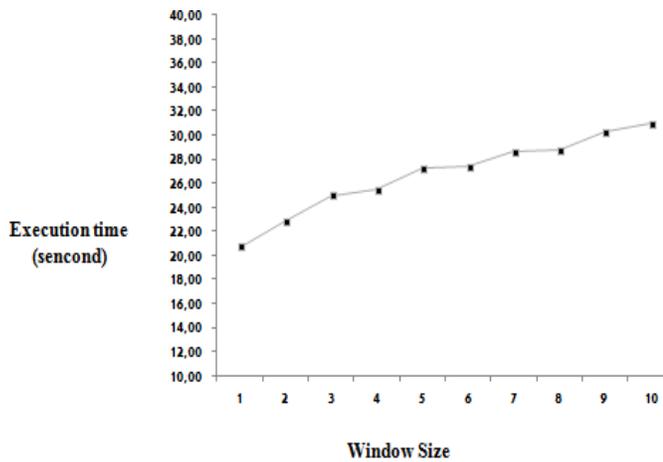


Fig. 12. Variation of the Execution Time Values with Different Window Sizes. EVS=50, MFW=2.

In contrast with the impact of the parameters EVS and WS discussed above, the execution time of the embedding layer decreases by increasing the MFW values in the used corpus. As shown in Fig. 13, the execution time can be reduced to half ($\approx 55,96\%$) of its initial value when MFW changes from MFW =1 to MFW = 10.

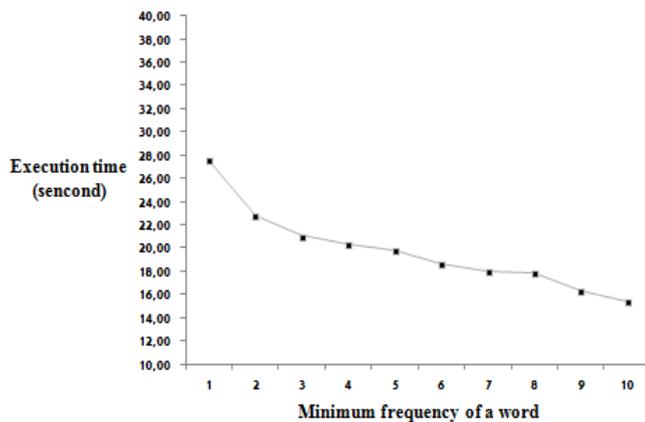


Fig. 13. Variation of the Execution Time Values with Different MFW Values. EVS=50, WS=2.

As a main result of these first experiments, the execution time of the used embedding layer can be improved if the set of parameters (WS, MFW, EVS) is optimized. Consequently,

TABLE V. VARIATION OF LOSS VALUES WITH DIFFERENT DEEP LEARNING MODELS AND EVS VALUES. WS =2, MFW=2

Models	Embedding Vector Size (EVS)				
	10	50	80	100	140
Simple GRU	1,16%	0,62%	0,53%	0,52%	0,49%
Bidirectional GRU	1,12%	0,46%	0,31%	0,27%	0,26%
Stacked GRU	1,11%	0,55%	0,55%	0,55%	0,48%
Simple LSTM	1,09%	0,63%	0,59%	0,55%	0,55%
Stacked LSTM	1,05%	0,57%	0,54%	0,53%	0,51%
Bidirectional LSTM	0,96%	0,44%	0,30%	0,27%	0,25%

this optimized execution time can positively impact the total duration of DLN computation process (embedding layer and DLN models execution times).

2) *The impact of embedding vector size (EVS) on fake news detection performances:* This experiment assesses the effects of FastText process and News title feature on Fake News detection performances by building the used six DLN architectures with less EVS sizes and then more EVS sizes. The used EVS ranges from 10 to 140 (140 \approx half of the maximum length of news titles). As shown in Table V, as EVS values increase, the loss values decrease slightly for all the used six DLN architectures. When EVS changes from EVS =10 (small size) to EVS = 140 (high-dimensionality), the approximate decreases in the loss are around 0.67% for simple GRU, 0.86% for bidirectional GRU, 0.63% for stacked GRU, 0.54% for simple LSTM, 0.54% for stacked LSTM, and 0.71% for bidirectional LSTM. According to the Table VI, the News Title feature has a significant influence on the accuracy for all used detection models. These six architectures can achieve an accuracy that exceeds 98% for small and high values of EVS. The accuracy value increases slightly by increasing EVS. If EVS go from 10 to 140, the approximate increases in accuracy are around 0.83% for simple GRU, 1,06% for bidirectional GRU, 0,76% for stacked GRU, 0,72% for simple LSTM, 0,71% for stacked LSTM, and 0,99% for bidirectional LSTM. Compared to the other used models, the bidirectional LSTM network achieved the best accuracy for all used EVS values.

The ROC curves plotted with different models and EVS values show that News Title feature and the six used architectures provide a high fake news detection accuracy (Fig. 14). These curves show that there is significant improvement in News classification accuracy with lower decision thresholds, in particular with the bidirectional LSTM network. The AUC (Area Under ROC Curve) ranges in value from 0.95 to 0.97 when the used EVS values range from 10 to 140.

As a main result of this experiment, it's possible to achieve better Fake News detection performance by choosing a low Embedding Vector Sizes (EVS).

TABLE VI. VARIATION OF ACCURACY VALUES WITH DIFFERENT DEEP LEARNING MODELS AND EVS VALUES. WS =2, MFW=2

Models	Embedding Vector Size (EVS)				
	10	50	80	100	140
Simple GRU	98,63%	99,30%	99,45%	99,45%	99,46%
Bidirectional GRU	98,68%	99,50%	99,69%	99,74%	99,74%
Simple LSTM	98,68%	99,31%	99,37%	99,40%	99,40%
Stacked GRU	98,74%	99,29%	99,33%	99,42%	99,50%
Stacked LSTM	98,75%	99,38%	99,44%	99,44%	99,46%
Bidirectional LSTM	98,82%	99,61%	99,73%	99,79%	99,81%

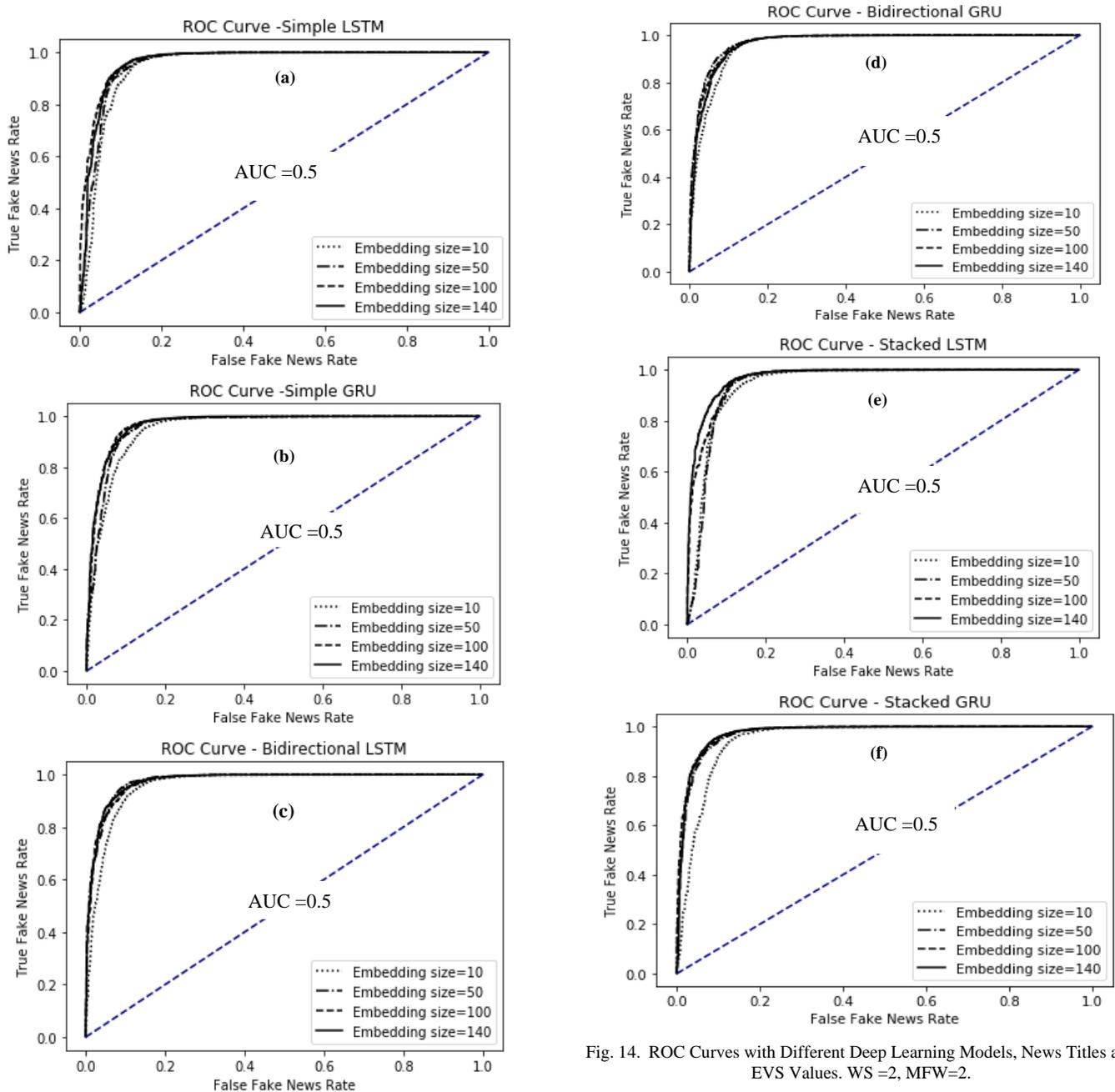


Fig. 14. ROC Curves with Different Deep Learning Models, News Titles and EVS Values. WS =2, MFW=2.

TABLE VII. VARIATION OF LOSS VALUES WITH DIFFERENT DEEP LEARNING MODELS AND WS VALUES. EVS =50, MFW=2

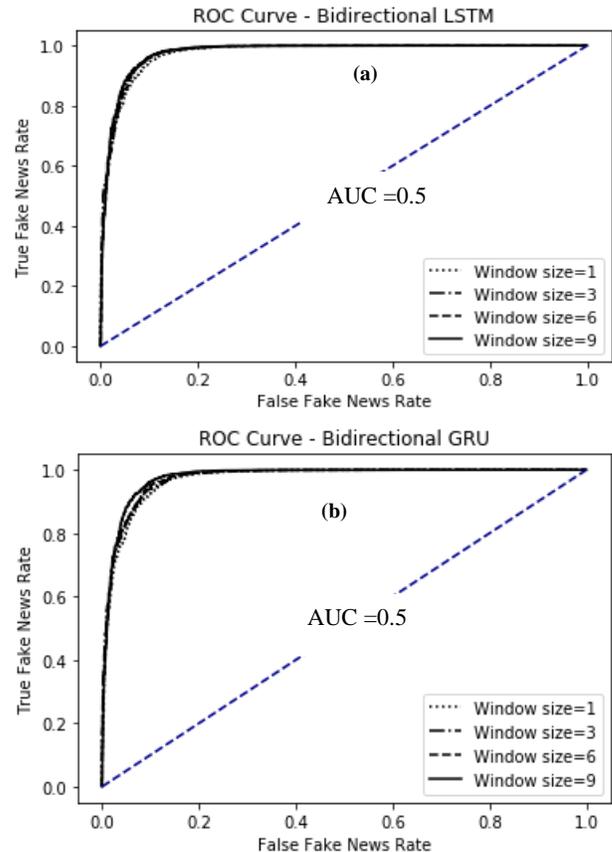
Window Size (WS)						
	1	3	5	7	9	10
Bidirectional LSTM	0,35%	0,28%	0,30%	0,31%	0,32%	0,29%
Bidirectional GRU	0,36%	0,29%	0,30%	0,31%	0,34%	0,31%
Simple GRU	0,61%	1,28%	0,84%	0,76%	0,64%	0,30%
Stacked GRU	0,62%	0,64%	0,68%	0,53%	0,65%	0,60%
Simple LSTM	0,62%	0,51%	0,55%	0,50%	0,52%	0,54%
Stacked LSTM	0,64%	0,55%	0,58%	0,51%	0,55%	0,57%

TABLE VIII. VARIATION OF ACCURACY VALUES WITH DIFFERENT DEEP LEARNING MODELS AND WS VALUES. EVS =50, MFW=2

Window Size (WS)						
	1	3	5	7	9	10
Bidirectional LSTM	99,32%	99,65%	99,01%	99,12%	99,27%	99,68%
Bidirectional GRU	99,61%	99,70%	99,69%	99,66%	99,63%	99,68%
Simple GRU	99,32%	99,29%	99,24%	99,41%	99,29%	99,31%
Stacked GRU	99,31%	99,44%	99,40%	99,44%	99,42%	99,41%
Simple LSTM	99,28%	99,39%	99,35%	99,45%	99,40%	99,36%
Stacked LSTM	99,61%	99,73%	99,67%	99,68%	99,67%	99,71%

3) *The impact of window size (WS) on fake news detection performances:* Generally, Window Size (WS) has the impact of giving more importance to closer words. Smaller WS lead to similar interchangeable words. Larger WS lead to similar related words. This experiment assesses the impact of WS by using less WS values and then more WS values. The used WS ranges from WS=1 to WS=10. According to the Table VII, increasing WS values causes a small variance of loss values. When WS changes from WS =1 (small Window Size) to WS = 10 (high Window Size), the difference between the minimum and the maximum of loss values does not exceed 0,07% for Bidirectional LSTM and Bidirectional GRU, 0,15% for simple LSTM, Stacked LSTM / Stacked GRU and 1% for simple GRU.

These architectures achieve high accuracy exceeds 98% for small and high used values of WS. If WS go from 1 to 10, the approximate difference between the minimum and the maximum of accuracy values does not exceed 0,09% for Bidirectional GRU, 0,17% for simple LSTM / simple GRU, 0,13% for Stacked LSTM / Stacked GRU and 0.67% for Bidirectional LSTM (Table VIII). By testing smaller and larger WS values, the ROC curves plotted with different Window Sizes show that news title feature and FastText process provide a high Fake News classification accuracy (Fig. 15).



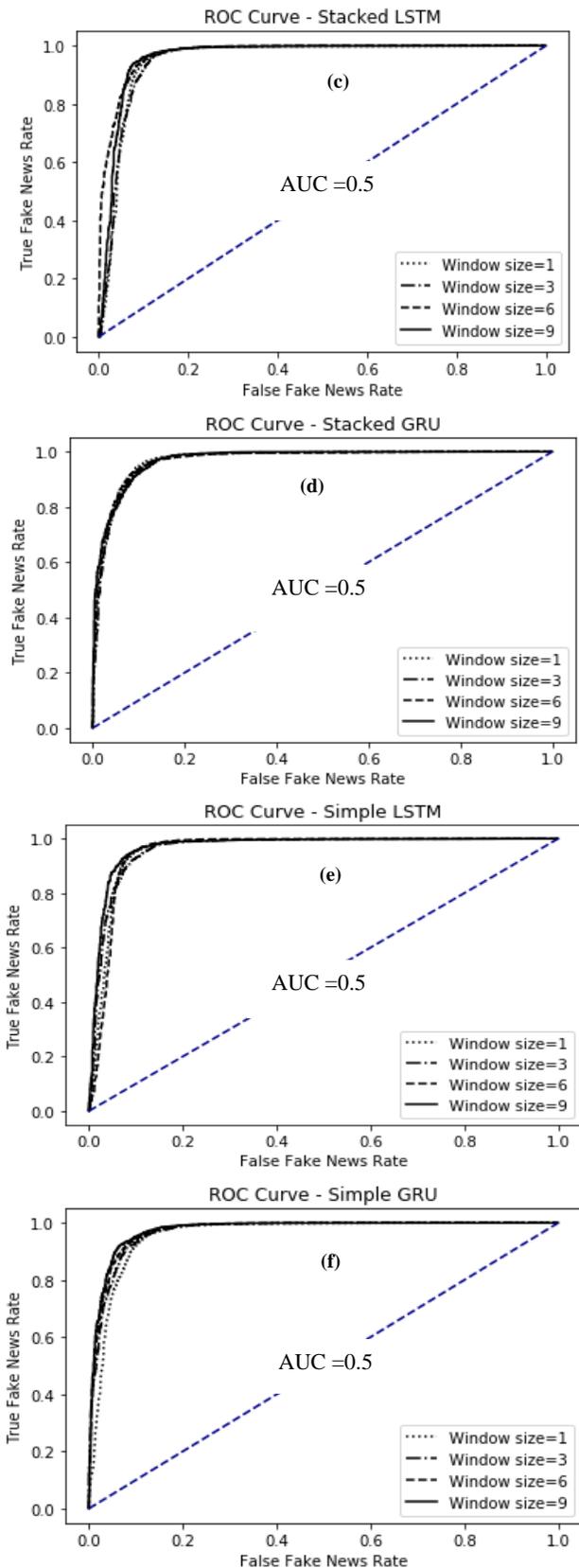


Fig. 15. ROC Curves with Different Deep Learning Models and WS Values. EVS =50, MFW=2.

Increasing WS values impacts slightly the ROC decision thresholds. Important improvement in Fake News classification is obtained from lower decision thresholds for all the six used models. When WS values range from 1 to 10, AUC ranges in value from 0.96 to 0.97 for LSTM architectures and from 0.96 to 0.98 for GRU architectures. As main result of these experiments, the use of News Title feature with low dimensions of Windows Size (WS between two and three) are enough to capture enough information with to detect Fake News with high performance.

4) *The impact of minimum frequency of a word (MFW) on fake news detection performances:* The hyper-parameter MFW (Minimum Frequency of Word) specifies the minimum frequency of a word in the News Titles corpus for which the word embedding will be generated. This last experiment assesses the impact of MFW by testing less MFW values and then more MFW values. The used MFW ranges from 1 to 9. As shown in the Table IX, the loss value is slightly decreased by increasing the values of MFW. If MFW go from 1 to 9, it decreases by 0,11% for Bidirectional LSTM, 0,16% for Bidirectional GRU, 0,04% for Stacked LSTM, 0,17% for Stacked GRU, 0,09% for Simple LSTM and 0,11% for Simple GRU. The minimum of loss values is observed for high values of MFW (MFW = 5 for simple GRU, MFW = 9 for all other models). The stacked GRU model shows the higher loss value for MFW between 1 and 5.

Compared to the other used models, Bidirectional LSTM and GRU models show the higher accuracy for all used MFW values. The ROC curves plotted with different MFW values show that news title feature used as input of FastText embedding layer provides a high Fake News classification accuracy for smaller and larger MFW values (Fig. 16). Increasing MFW value impacts slightly the ROC decision thresholds. Important improvement in the Fake News classification is obtained from lower decision thresholds for all the used models.

For higher values of MFW (MFW between 7 and 9), this maximum value is obtained by the simple GRU model.

The six used architectures achieve high accuracy exceeds 98% for small and high used values of MFW. According to the Table X, news title feature has a significant influence on the accuracy for all the used MFW values. As MFW values increase, the accuracy values increase slightly for all the used models. When MFW changes from MFW =1 (small frequency) to MFW = 9 (high frequency), the approximate increases in accuracy are around 0.14% for Bidirectional LSTM, 0.17% for Bidirectional GRU, 0.12% for Stacked / simple LSTM, 0.16% for Simple GRU, and 0.25% for Stacked GRU.

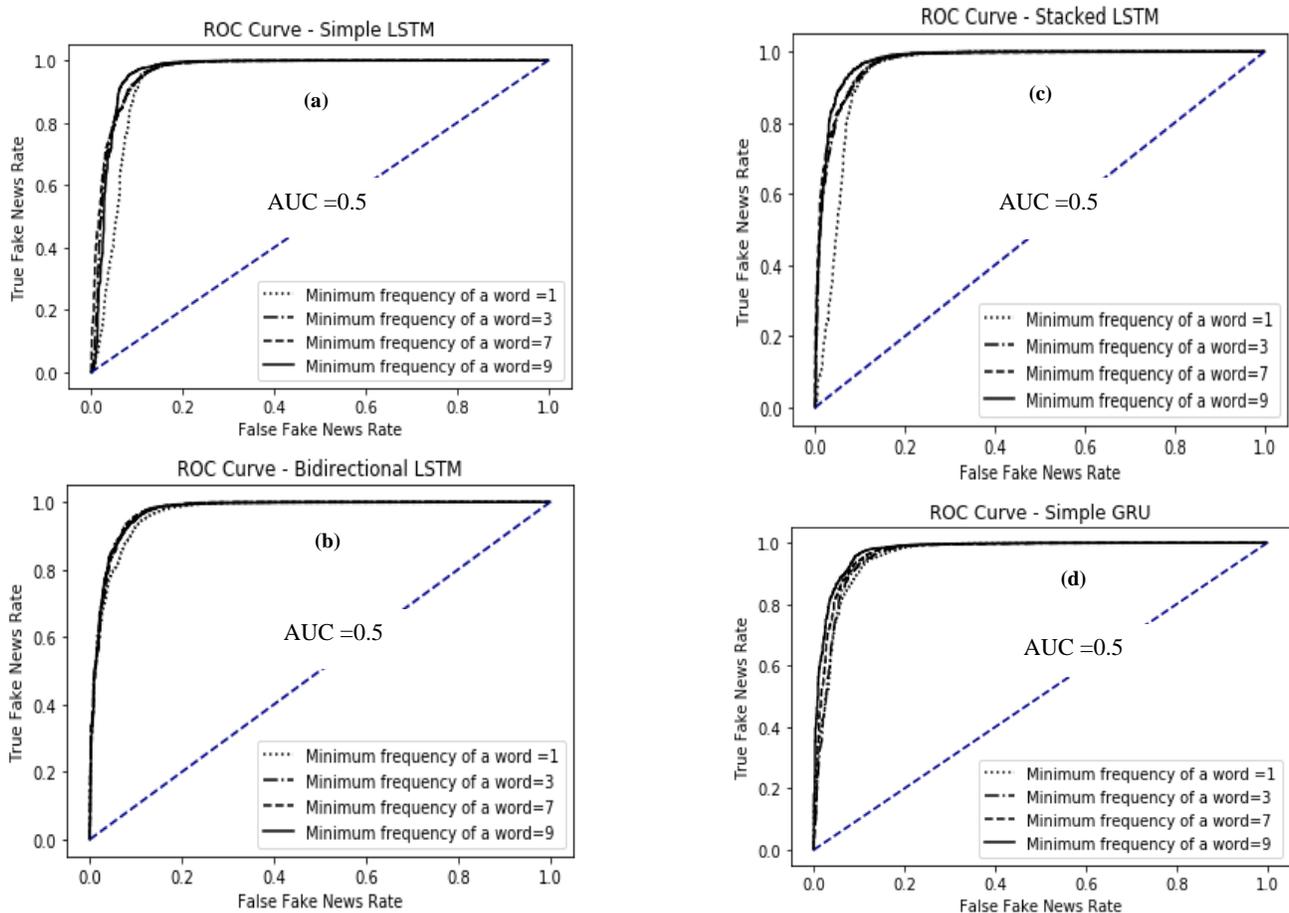
When MFW values range from 1 to 9, the AUC (Area Under the ROC Curve) ranges in value from 0.96 to 0.97 for all used LSTM architectures and from 0.96 to 0.98 for all used GRU architectures. The findings of this last experiment showed that News titles with low size of the minimum frequency of words (MFW) didn't affect significantly the performances of Fake News detection.

TABLE IX. VARIATION OF LOSS VALUES WITH DIFFERENT DEEP LEARNING MODELS AND MFW VALUES. EVS =50, WS=2

Models	Minimum frequency of a word (MFW)				
	1	3	5	7	9
Bidirectional LSTM	0,46%	0,39%	0,39%	0,39%	0,35%
Bidirectional GRU	0,47%	0,44%	0,42%	0,39%	0,31%
Stacked LSTM	0,60%	0,60%	0,58%	0,56%	0,56%
Simple LSTM	0,65%	0,59%	0,58%	0,58%	0,56%
Simple GRU	0,72%	0,66%	0,58%	0,60%	0,61%
Stacked GRU	0,75%	0,67%	0,64%	0,58%	0,58%

TABLE X. VARIATION OF LOSS VALUES WITH DIFFERENT DEEP LEARNING MODELS AND MFW VALUES. EVS =50, WS=2

Models	Minimum frequency of a word (MFW)				
	1	3	5	7	9
Bidirectional LSTM	99,56%	99,62%	99,65%	99,70%	99,70%
Bidirectional GRU	99,51%	99,55%	99,56%	99,59%	99,68%
Stacked LSTM	99,26%	99,32%	99,35%	99,39%	99,38%
Simple LSTM	99,29%	99,33%	99,35%	99,38%	99,41%
Simple GRU	99,20%	99,23%	99,33%	99,35%	99,36%
Stacked GRU	99,13%	99,25%	99,30%	99,37%	99,38%



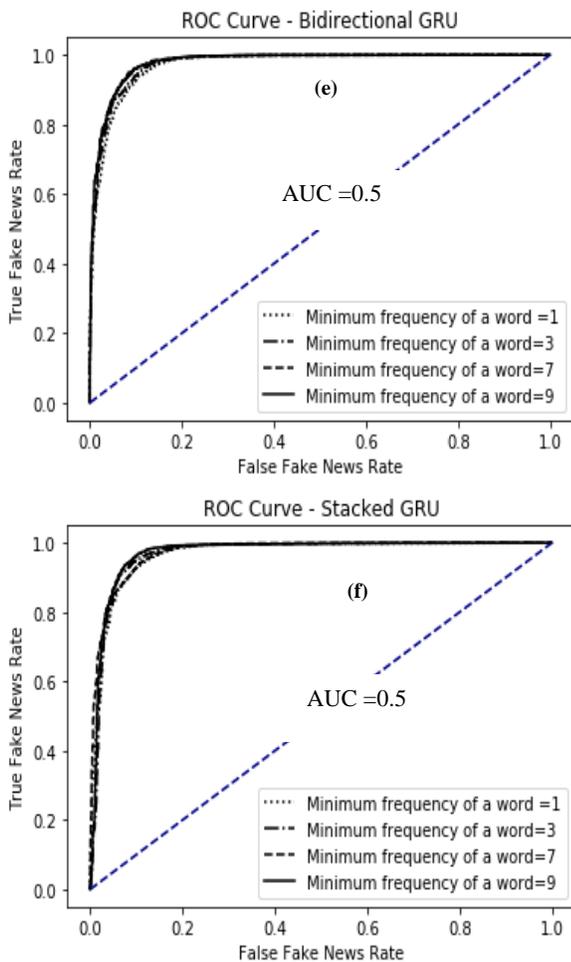


Fig. 16. ROC Curves with Different Deep Learning Models and MFW Values. EVS =50, WS=2.

IV. CONCLUSION

Today, Fake news phenomena are one of the world's top global risks. Therefore, social, political and economic environments need to be well-equipped to decipher and detect this massive, instantaneous and heterogeneous daily misinformation.

One of the promising fields of automated systems used to deal with this problem is Artificial Intelligence (AI), especially Deep Learning Networks (DLN).

In this context, the present paper focuses on the improvement of DLN Fake News detection by using Inverted Pyramid Pattern (IPP), and integrating the FastText process in many DLN architectures: Simple LSTM, Stacked LSTM, Bidirectional LSTM, Simple GRU, Stacked GRU and Bidirectional GRU. More precisely, this empirical study focuses on how news title feature and FastText embedding process can impact and improve the existing automatic DLN Fake News detection.

By testing these six architectures with many ranges of embedding layer main hyper-parameters (EVS, WS, MFW), these experiments showed that the news title feature and FastText process can have a significant improvement on DLN

automatic Fake News detection with accuracy rates exceeding 98%.

This paper is the first step of our future Fake News framework project based on many DLN architectures.

ACKNOWLEDGMENT

I would like to express my special thanks of gratitude to Khalid AHAJI, Hassan ELHANI, Fouad MOUSSAOUI, Abdelmoutalib MOUSSAOUI who encouraged and helped us in doing a lot of research.

Any attempt at any level can't be satisfactorily completed without the support and guidance of my parents, sisters and friends.

REFERENCES

- [1] Á. Figueira and L. Oliveira, "The current state of fake news: Challenges and opportunities," *Procedia Computer Science*, vol. 121, pp. 817–825, 2017.
- [2] C. Sindermann, A. Cooper, and C. Montag, "A short review on susceptibility to falling for fake political news," *Current Opinion in Psychology*, vol. 36, pp. 44–48, 2020.
- [3] D. D. Parsons, "The impact of fake news on company value: Evidence from Tesla and Galena Biopharma," [Online]. Available: https://trace.tennessee.edu/utk_chanhonoproj/2328/. [Accessed: 28-Nov-2021].
- [4] M. A. Rivera, "Fake news and hospitality research," *International Journal of Hospitality Management*, vol. 85, p. 102473, 2020.
- [5] F. A. Ozbay and B. Alatas, "Fake news detection within online social media using supervised artificial intelligence algorithms," *Physica A: Statistical Mechanics and its Applications*, vol. 540, p. 123174, 2020.
- [6] D. S and B. Chitturi, "Deep neural approach to fake-news identification," *Procedia Computer Science*, vol. 167, pp. 2236–2243, 2020.
- [7] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020.
- [8] Y. Ma and J. C. Principe, "A taxonomy for Neural Memory Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 6, pp. 1780–1793, 2020.
- [9] Deep learning crash course for beginners with python: Theory and applications of artificial neural networks, CNN, RNN, LSTM and autoencoders using tensorflow 2: Contains exercises with solutions and hands-on projects. AI Publishing, 2020.
- [10] P. Bahad, P. Saxena, and R. Kamal, "Fake news detection using bi-directional LSTM-recurrent neural network," *Procedia Computer Science*, vol. 165, pp. 74–82, 2019.
- [11] S. Al-Abri, T. X. Lin, M. Tao, and F. Zhang, "A derivative-free optimization method with application to functions with exploding and vanishing gradients," *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 587–592, 2021.
- [12] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
- [13] G. Tripepi, K. J. Jager, F. W. Dekker, and C. Zoccali, "Diagnostic methods 2: Receiver operating characteristic (ROC) curves," *Kidney International*, vol. 76, no. 3, pp. 252–256, 2009.
- [14] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [15] K. Englmeier, "The role of text mining in mitigating the threats from fake news and misinformation in times of Corona," *Procedia Computer Science*, vol. 181, pp. 149–156, 2021.
- [16] S. R. Sahoo and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using Deep Learning," *Applied Soft Computing*, vol. 100, p. 106983, 2021.

- [17] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "FNDNet – a deep convolutional neural network for fake news detection," *Cognitive Systems Research*, vol. 61, pp. 32–44, 2020.
- [18] D. S. and B. Chitturi, "Deep neural approach to fake-news identification," *Procedia Computer Science*, vol. 167, pp. 2236–2243, 2020.
- [19] M. H. Goldani, S. Momtazi, and R. Safabakhsh, "Detecting fake news with Capsule Neural Networks," *Applied Soft Computing*, vol. 101, p. 106991, 2021.
- [20] M. Farouk, "Measuring text similarity based on structure and word embedding," *Cognitive Systems Research*, vol. 63, pp. 1–10, 2020.
- [21] D. Jatnika, M. A. Bijaksana, and A. A. Suryani, "Word2Vec model analysis for semantic similarities in English words," *Procedia Computer Science*, vol. 157, pp. 160–167, 2019.
- [22] J. Choi and S.-W. Lee, "Improving FastText with inverse document frequency of subwords," *Pattern Recognition Letters*, vol. 133, pp. 165–172, 2020.
- [23] A. Toxboe, "Article list design pattern," *UI Patterns*. [Online]. Available: <http://ui-patterns.com/patterns/ArticleList>. [Accessed: 30-Nov-2021].
- [24] C. Bisailon, "Fake and real news dataset," *Kaggle*, 26-Mar-2020. [Online]. Available: <https://www.kaggle.com/clmentbisailon/fake-and-real-news-dataset>. [Accessed: 30-Nov-2021].
- [25] P. S. Janardhanan, "Project Repositories for Machine Learning with tensorflow," *Procedia Computer Science*, vol. 171, pp. 188–196, 2020.
- [26] V.-H. Nhu, N.-D. Hoang, H. Nguyen, P. T. Ngo, T. Thanh Bui, P. V. Hoa, P. Samui, and D. Tien Bui, "Effectiveness assessment of keras based deep learning with different robust optimization algorithms for shallow landslide susceptibility mapping at tropical area," *CATENA*, vol. 188, p. 104458, 2020.
- [27] Y. Kwon, S. Lee, H. Yi, D. Kwon, S. Yang, B.-gon Chun, L. Huang, P. Maniatis, M. Naik, and Y. Paek, "Mantis: Efficient predictions of execution time, energy usage, memory usage and network usage on smart mobile devices," *IEEE Transactions on Mobile Computing*, vol. 14, no. 10, pp. 2059–2072, 2015.