CTU Journal of Innovation and Sustainable Development
ISSN 2615-9422 | e-ISSN 2815-5602

DOI:10.22144/ctujoisd.2023.033

Detecting Vietnamese fake news

Vo Duc Vinh[1*], and Do Phuc[2]
[1]*VNPT - Dong Nai, Viet Nam*
[2]*University of Information Technology, Viet Nam National University Ho Chi Minh City, Viet Nam*
**Corresponding author (vdvinh.dni@gmail.com)*

Article info.

Received 28 Jul 2023
Revised 10 Sep 2023
Accepted 23 Sep 2023

Keywords

Factual news, fake news, false news, machine learning, neural network, true news

ABSTRACT

This paper focuses on constructing a dataset consisting of both fake news and factual news in the Vietnamese language. We employ Deep Learning models, namely Long Short-Term Memory, bidirectional Long Short-Term Memory, and Convolutional Neural Network - bidirectional Long Short-Term Memory, to identify Vietnamese fake news. The performance evaluation of the models includes assessing the prediction ratio Area Under The Curve of each model and providing insights into their computational efficiency. Additionally, these three models evaluate the contribution of deep learning techniques for fake news detection and emphasize the potential for exploring interconnections between neural networks in addressing automatic Vietnamese fake news detection.

1. INTRODUCTION

With the advancements in science and technology, particularly in the realm of Information and Communication Technology, a novel method of exchanging information has emerged online information exchange through social networks. Currently, social networks are extensively utilized by many individuals because of their ability to rapidly disseminate information and reach a larger audience compared to traditional forms of communication. Social networks facilitate easy sharing, commenting, and discussions among users and their friends. However, the internet, particularly social networking sites, has witnessed the proliferation of fake accounts that disseminate unverified information on various topics such as politics, epidemics, natural disasters, climate change, hydrological phenomena, superstitions, and false advertising. This phenomenon has led to confusion, disruption, and a significant impact on people's daily lives. Many point to the 2016 U.S. presidential election campaign as having been influenced by fake news. These misleading articles typically employ sensational content to captivate viewers' attention, amplify the influence of the posters on social networks, and generate increased traffic for financial gain.

Currently, the detection of fake news in Viet Nam relies entirely on manual human intervention. Only a few studies conducted by domestic researchers have addressed fake news detection. One study by Nguyen and Nguyen (2020) proposes an approach to detect fake news on social networking sites (SNS) using linguistic features synthesized by PhoBERT. Another article introduces the tasks shared on ReINTEL, including three phases: launch, public testing, and private testing, all related to fake news (Le et al., 2020). However, these findings are still at a preliminary stage and not yet applicable.

Internationally, fake news research often intersects with terms and concepts related to fake news (Allcott & Gentzkow, 2017; Vosoughi et al., 2018). The challenges in this research domain begin with defining what constitutes fake news. Currently, there is no universally accepted definition; however, it is commonly understood as "an intentionally false article that is difficult to verify" (Allcott & Gentzkow, 2017; Shu et al., 2017).

Some researchers have proposed the utilization of neural network models and deep learning techniques for fake news detection, yielding promising results (Zhou & Zafarani, 2020). Recent efforts involve combining multiple models and methods to enhance the accuracy of fake news detection. Combining CNN-RNN models showed success in various classification and regression tasks, as they can capture both local and sequential features of the input data. For instance, these models have been employed in emotion detection (Kollias & Zafeiriou, 2020) or feature extraction through model fusion (Elhadad et al., 2020).

In the international sphere, there have been research endeavors and solutions employing deep learning models for fake news detection, yielding favorable outcomes. However, the research on detecting fake news specifically in the Vietnamese language remains limited.

Recognizing and detecting fake news in Vietnamese encounters various difficulties and challenges. First, constructing a reliable fake news recognition system requires a substantial, accurate, and up-to-date dataset encompassing both real and fake news. Given the vast volume of information generated and disseminated on the internet through diverse methods, this becomes a critical requirement. Second, Vietnamese sentence syntax differs from English. For demonstration, the sentence "Ông già đi nhanh quá" (the old man goes too fast) can be understood as 'Ông già' + 'đi' + 'nhanh quá' (The old man is walking too fast) or 'Ông' + 'già đi' + 'nhanh quá' (The old man grows old too fast)' in English. This disparity poses a challenge in accurately interpreting Vietnamese sentences. Third, determining the credibility of news sources proves to be exceedingly challenging. With the rapid development of diverse social networking sites, keeping official websites updated and reliable becomes arduous.

In summary, building a system for recognizing real and fake news in Vietnamese presents numerous challenges within the realm of scientific research.

This article contributes to the field in the following ways:

– Collection and creation of Vietnamese datasets for fake news detection.

– Researching three deep learning machine models (LSTM, Bi-LSTM, CNN-BiLSTM) for automatic detection of Vietnamese fake news.

– Conducting experiments and analyzing the results using deep learning models on Vietnamese datasets for detecting fake news in Vietnamese.

The subsequent sections of the paper are organized as follows: 2) Methods for automatic fake news detection 3) Our proposed method 4) Conclusion and future works.

2. METHODS FOR AUTOMATIC FAKE NEWS DETECTION

2.1. Deep learning techniques for Fake News Detection

2.1.1. Long Short-Term Memory (LSTM) Network

LSTM (Long Short-Term Memory) is an improved variant of RNN designed to address the issue of retaining information over long sequences. LSTMs are inherently capable of remembering information for extended periods without specific training. Each node in an LSTM can store information without external intervention.

2.1.2. Bi-LSTM Network Model

Bi-LSTM is a type of recurrent neural network commonly used in natural language processing. It processes input in both the forward and backward directions, allowing it to leverage information from both ends. It effectively models sequential dependencies between words and phrases in both directions of the sequence. Bi-LSTM adds an additional layer of LSTM that processes the input sequence in reverse. The outputs from both the forward and reverse LSTM layers can then be combined using operations such as averaging, summation, or multiplication.

2.1.3. Combined CNN and Bi-LSTM Network Model

The combined model of CNN (Convolutional Neural Network) and Bi-LSTM is widely used in processing string data, particularly in natural language processing and text classification. This model leverages the spatial feature learning capabilities of CNNs and the sequence relationship learning capabilities of Bi-LSTM.

In the combined model, each network serves a distinct function:

1) Convolutional Neural Network (CNN):

– Function: The CNN network is used to learn spatial features from the input data.

– Operation: The CNN network applies filters and kernel sizes to extract spatial information from the data. By sliding filters over word representations or word sequences, CNNs can identify significant features. These features are then abstracted through techniques like MaxPooling to reduce the dimension of the representation.

2) Bidirectional LSTM (BiLSTM):

– Function: The BiLSTM network is used to learn sequential relationships within the data.

– Operation: The BiLSTM network learns information from both directions of the input sequence. With the LSTM architecture, it can store information in long-term memory and selectively "forget" irrelevant information. By running two LSTMs in parallel, one in the forward direction and the other in the reverse direction, BiLSTM can capture complex sequential dependencies and understand contextual information from both ends of a sentence or text.

2.1.4. Word-based tokenization

This tokenization method is widely employed in the domain of text analysis. It serves to partition a given textual input into discrete units, primarily words, facilitating the extraction of textual semantics through the examination of word sequences. For the effective utilization of natural language processing (NLP) models, language-specific word segmentation techniques are employed. It is important to note that, depending on the specific analytical task at hand, a given text may undergo distinct tokenization schemes. Each resultant token typically possesses a distinctive identity, represented by a unique identifier (ID), thus serving as a method of digital encoding and token identification.

3. OUR PROPSED METHOD

Research in fake news detection has explored the application of deep learning models and combined neural networks to address this problem. However, most of these studies have focused on English datasets, and implementing Vietnamese datasets is still limited.

Table 1. Table comparing pros and cons of 3 models (Kavianpour et al., 2021)

No.	Model	Advantages	Disadvantages
1	LSTM	- LSTM is capable of long-term data processing and identifying complex relationships in sequence data. - LSTM has the ability to handle long dependencies between components in the sequence data. - LSTM has the ability to learn the spatial and temporal characteristics of sequence data.	- The architecture of the Long Short-Term Memory (LSTM) network is intricate, characterized by a significant quantity of parameters. Consequently, a substantial allocation of computational resources is imperative for both the training and prediction phases. - LSTM is quite sensitive to the phenomenon of information loss (vanishing/exploding gradients) during training. - LSTM is not efficient in handling spatial data, such as images.
2	Bi-LSTM	- Bi-LSTM combines both forward and reverse processing in sequence data processing, improving the ability to understand and model dependency relationships in the data. - Bi-LSTM can learn features from both the past and future of a location in the sequence data.	- Bi-LSTM doubles the number of parameters compared to conventional LSTM, thus requiring more computational resources for training and prediction. - Bi-LSTM may have difficulty processing data represented by matrices, such as images, because it cannot exploit the spatial structure of the data.
3	CNN-BiLSTM	- CNN-BiLSTM combines both the spatial feature learning of CNNs and the ability to understand and model dependency relationships in sequence data of Bi-LSTM. - CNN-BiLSTM can discover features at different levels of abstraction in sequence data, helping to improve model performance.	- The CNN-BiLSTM model exhibits a sophisticated arrangement and a notable abundance of parameters, thereby demanding considerable computational resources for the training and prediction processes. - Efficient acquisition of spatial and temporal features in CNN-BiLSTM models could necessitate a substantial volume of training data.

In this paper, we have utilized a deep learning model, specifically Deep Learning, to evaluate three neural network models: LSTM, Bi_LSTM, and CNN-BiLSTM networks. The selection of these models was based on their respective advantages. Table 1 (Kavianpour et al., 2021) provides a comparative analysis of these models, highlighting their strengths and characteristics.

3.1. Data sources

For validation and training of the models, we collect and perform testing under two mixed datasets:

3.1.1. VNDF-Vietnamese-fake-news-datasets

VNDF-Vietnamese-fake-news-datasets downloaded at the link (https://dagshub.com/thanhhocse96/VFND-vietnamese-fake-news-datasets), which is Vietnamese false news dataset archived between 2017 and 2019. This news are categorized as fake based by several sources, cross-referencing sources cited or classified by the community.

To effectively detect false information, it is essential to establish a clear definition of fake news during the process of generating datasets. Various aspects should be considered, including misinformation, deceptive content, and fabricated narratives. The following steps outline the procedure for constructing datasets containing both genuine and fake news (Figure 1):

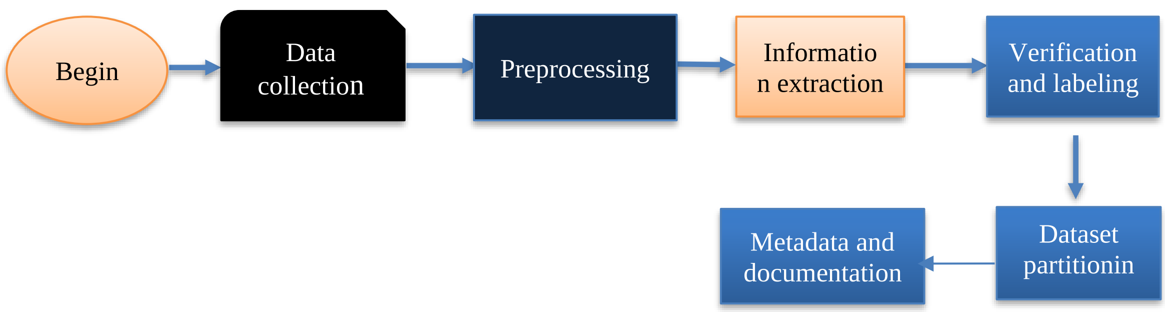

Figure 1. The process of building a Vietnamese dataset

Step 1 (Data collection): Gather a collection of credible electronic articles as a source of authentic information, (such as: VnExpress, Vietnamnet, Thanh Nien, Tuoi Tre, and Online News). Additionally, compile a set of counterfeit electronic articles to generate fake news data, (such as: Đại kỷ Nguyên “Great Yuan”, SOHA, Đài Á Châu Tự Do “Radio Free Asia”, RFI tiếng Việt “Vietnamese RFI”, and Fanpage). Ensure that the selected articles have substantial viewership, comments, and shares.

Step 2 (Data preprocessing): After obtaining the data from social media platforms in a predefined structure, preprocess the data by converting it to the correct format and removing special characters, redundant Vietnamese words, and non-Vietnamese words. For instance, eliminate special characters (such as : “@”, “%”, “$”, as well as redundant phrases like "a ha", "ê hê", “ai ai”).

Step 3 (Information extraction): Extract the most significant information from each document, condensing it into a concise version that retains the core details of the original article. Ensure that the grammar and spelling remain accurate. Employ the automatic Vietnamese text extraction method developed by Dinh Truong Quoc and Dung Nguyen Quang (2012) (Quoc & Dung, 2012) to extract important sentences from the collected test dataset sourced from social networks and real news datasets obtained from trusted websites.

For example: a sentence like "Trận đấu kết thúc với chiến thắng đầy nỗ lực của đội tuyển Việt Nam. Cầu thủ số 10 đã ghi bàn thắng quyết định và mang về chiến thắng cho đội nhà." (The match ended with a victory full of efforts of the Vietnamese team. Player No. 10 scored the decisive goal and brought the victory to the home team.) would be extracted into two sentences: "Trận đấu kết thúc với chiến thắng đầy nỗ lực của đội tuyển Việt Nam." (The match ended with a victory full of efforts of the Vietnamese team.) and "Cầu thủ số 10 đã ghi bàn thắng quyết định và mang về chiến thắng cho đội nhà."(Player No. 10 scored the decisive goal and brought the home team victory.).

Step 4 (Verification and labeling): Establish principles for determining similarity, specifically focusing on the semantic similarity between important sentences in the test dataset and the dataset containing real news. Assemble a team of collaborators who adhere to the predefined rules to examine and label the articles as either true or false.

For example: sentences like "Hôm nay trời đẹp" ("The weather is nice today") and "Thời tiết hôm nay rất đẹp" (Very nice weather today) would be deemed semantically similar.

Step 5 (Dataset partitioning): Divide the labeled dataset into training, validation, and testing sets, while ensuring a balanced distribution of genuine and fake articles across the sets to avoid any biases. If the dataset exhibits specific characteristics or variations, consider employing stratified sampling techniques.

Step 6 (Metadata and documentation): Include relevant metadata for each article, such as its source, publication date, author, and any other pertinent information. Additionally, create comprehensive documentation that describes the dataset's composition, potential biases, limitations, and any other relevant details.

For example: the news article titled "How do Go-Jek and Grab win the Indonesian market?" extracted from the newspaper "news.zing.vn" on 19/7/2017.

3.1.2. True and fake news

True and Fake news downloaded at the link (https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset) and were translated to Vietnamese and went under pre-processing.

The data were trained as follows (Figure 2):

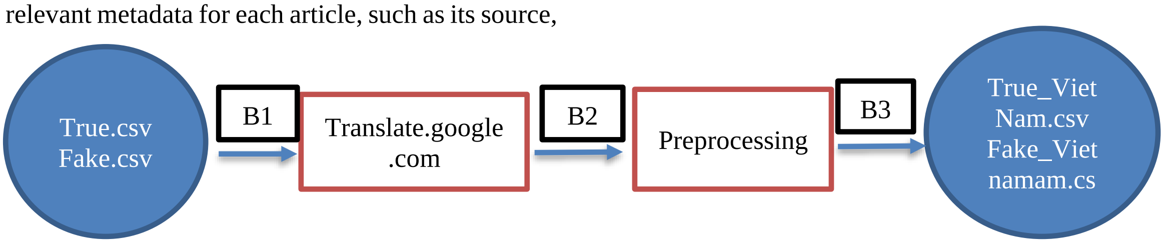

Figure 2. The process of processing English-language dataset into Vietnamese

– Step 1: Translate.google.com is utilized to translate into Vietnamese.

– Step 2: After translating into Vietnamese, the data set is checked for font errors, incorrect accents, spelling, and special characters are removed.

– Step 3: After preprocessing the data sets are saved 2 data files True_VietNam.csv and Fake_VietNam.csv.

After normalizing the dataset, the data tables are saved by name and structure as follows:

+ The truth data set is stored with the file name True_VietNam_CSV.csv, including columns of information (title, text, subject, date) with a total of 21527 messages as shown in Table 2.

Table 2. Fact dataset

Title	Text	Subject	Date
Many planes cannot land at Tan Son Nhat	It rained heavily in Saigon, trees fell and people died On the evening of November 25, the circulation of Typhoon Usagi caused heavy rain throughout Ho Chi Minh City and the Southern provinces, in many places over 200 mm. A series of flights of Vietnam Airlines, Vietjet Air, Jetstar Pacific departing from Tan Son Nhat airport announced delays due to bad weather. Flights to Ho Chi Minh City are also behind schedule, many planes even have to fly around to land at Tan Son Nhat airport, or land at another airport. Passengers on flight TG556 from Bangkok to Tan Son Nhat could not board so they had to fly back. Photo: Ha Nguyen. Many international flights returning to Tan Son Nhat had to fly many times to try to land but failed. Including the Bangkok - HCM flight number TG556, after nearly 40 minutes of roundabout flight, had to return to Thailand.	vnexpress.net	25/10/2020

+ The false news data set is stored under the name Fake_VietNam_CSV.csv, which includes titles, texts, subjects, and dates, with a total of 22475 news segments as in Table 3.

Table 3. Fake dataset

Title	Text	Subject	Date
The Japanese Prime Minister bowed his head and apologized for the unsportsmanlike spirit of the team	On the morning of June 19, Japanese Prime Minister Shinzo Abe publicly apologized to the Emperor and all people for the unsportsmanlike playing spirit of the Japanese team at the 2018 World Cup. With the spirit of Samurai warriors, the Japanese team has acted without martial spirit in competing at the World Cup, leaving a lot of criticism and dissatisfaction among the entire people, ruining the image of steadfastness." strength of the Japanese people in the international arena. As the leader, I sincerely accept responsibility and deeply apologize to the people," Mr. Abe bowed his head to accept responsibility. I admire this action of the Japanese Prime Minister. He is very conscious of national image and honor! Japan is respected and admired by the world not because it is an economic powerhouse, but because it is a powerhouse of self-esteem!	www.ipick.vn	19/06/2018

+ The data set "stopword" downloaded at the link (https://github.com/stopwords/vietnamese-stopwords/blob/master/vietnamese-stopwords.txt) and stored under the name stopwords_vietnam.txt, including stopwords with a total number of 1942 (Example stopwords as: a lô, a ha, ai ai…).

3.2. Model training procedure

After collecting and preprocessing the Vietnamese dataset, including checking for special characters, misspelled words, and font errors, the model is trained online using the Colab environment with the following specifications: RAM: 13GB, GPU,

Disk: 78GB. The parameters used for training models Table 4:

Table 4. The parameters used for training models

LSTM	Bi-LSTM	CNN-BiLSTM
embedding_vector_features=40 Dropout(0.3) LSTM(100) Dense(1,activation='sigmoid') loss='binary_crossentropy' optimizer='adam' metrics=['accuracy']	embedding_vector_features=40 Dropout(0.3) Bidirectional(LSTM(100)) Dense(1,activation='sigmoid') loss='binary_crossentropy' optimizer='adam' metrics=['accuracy']	embedding_vector_features=40 Dropout(0.3) Conv1D(32, 5, activation='relu') Bidirectional(LSTM(100)) Dense(1,activation='sigmoid') loss='binary_crossentropy' optimizer='adam' metrics=['accuracy']

Note:

Embedding: Embedding is a parameter used to represent words or sentences as lower-dimensional vectors, which helps improve computational efficiency and reduce data dimensionality. The syntax for embedding is embedding (batch_size, sequence_length, embedding_dimension), where batch_size refers to the number of samples in each batch (can be dynamically changed), sequence_length is the length of each sequence, and embedding_dimension is the length of the vector representing each word.

Dropout: Dropout is a parameter used to reduce overfitting and improve the training speed of the model. The syntax for dropout is dropout(batch_size, sequence_length, embedding_dimension).

Dense: Dense represents a fully connected layer, where each node in the layer is connected to all nodes in the previous layer. The Dense class takes the input information and converts it into predictive output. The syntax for Dense is Dense (None, x), where None indicates that the number of inputs for the Dense layer will be automatically determined based on the previous layer's inputs, and x is the desired output number of nodes in the Dense layer.

Activation('sigmoid'): Activation is a nonlinear activation function that transforms the input of a processing unit into the corresponding output. In this case, 'sigmoid' activation function is used, which outputs values between 0 and 1, representing probabilities. It helps the neural network learn and represent nonlinear relationships between input and output features.

Loss ('binary_crossentropy'): Loss is a function that measures the difference between the predicted output and the actual value, allowing the model to optimize itself in binary classification problems. 'binary_crossentropy' is the specific loss function used.

Bidirectional(None, x): Bidirectional is used to create a bidirectional LSTM layer. The syntax is Bidirectional (None, x), where None indicates that the batch size can be flexible, and x represents the number of units in the bidirectional LSTM layer. A higher number of units allows the model to learn and represent more complex relationships in sequence data.

The model training steps are outlined in Figure 3.

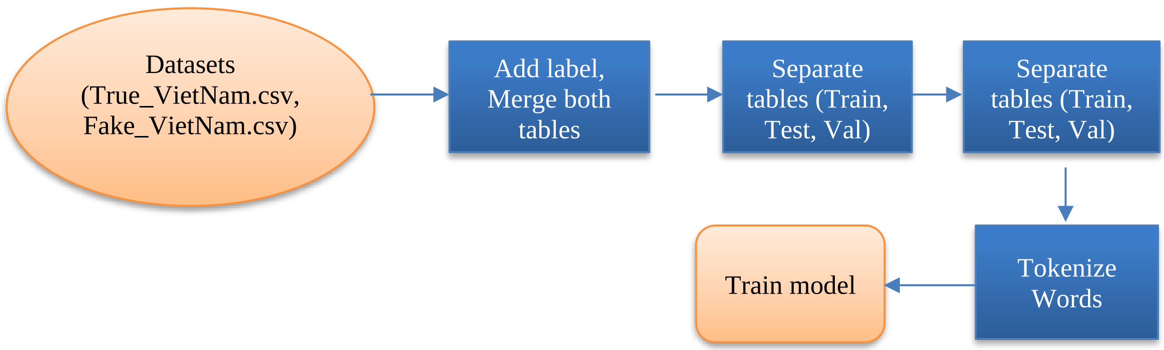

Figure 3. Model training steps

Step 1: Assign the label "1" to the dataset "True_VietNam_CSV.csv" and label "0" to the dataset "Fake_VietNam_CSV.csv". Merge the two datasets into one.

Step 2: Divide the combined data set in step 1 (44002 records) into data sets with the following ratios:

– Training data set (Train) with 28162 records (64% of total data set)

– Test data set (Test) with 8800 records (20% of total data set)

– Test data set (Val) with 7040 records (16% of total data set)

Step 3: Perform stopword removal to eliminate commonly used words that do not carry significant meaning.

Step 4: Tokenize the text by separating words and encode the words to numerical representations.

Step 5: Build and train the model.

3.3. Experimental Results

To assess the model's accuracy, the following indicators are considered:

– Precision: It measures the accuracy (percentage) of correct predictions among all the true and false predictions made by the model.

– Recall: It determines the percentage of correct predictions among all the actual instances of true or false news.

– Accuracy: It represents the percentage of correct predictions made by the model among all the true and false instances in the test dataset. A higher accuracy indicates a more accurate prediction model.

After testing the LSTM, Bi-LSTM, and CNN-BiLSTM models, the AUC percentages of the three models show minimal differences. However, the CNN-BiLSTM model yields the best results. The table below evaluates the measurement indicators for the three models:

In this section, we will provide the evaluation results for the LSTM model. The indices for the remaining models are similar.

For the LSTM model:

– Precision = 0.95: This indicates that 95% of the predictions made by the model for fake news are correct among all the predictions for both real and fake news.

– Recall = 0.97: This means that the model correctly predicts 97% of the fake news instances among all the actual instances of both fake and real news.

– F1-score = 0.96: The F1-score is the average of the precision and recall, which in this case is 0.96. It represents a balanced measure of the model's accuracy in predicting both fake and real news.

AUC = 0.959: The AUC (Area Under the Curve) is a percentage that indicates the model's ability to correctly predict fake news among all the true and fake news instances in the test dataset. The higher the AUC, the more accurate the model. In this case, the AUC is 0.959.

Table 5. Table of results for the 3 models

Models	Label	Precision	Recall	F1-score	Support	AUC
LSTM	0	0.95	0.97	0.96	4471	0.959
	1	0.96	0.95	0.96	4331	
Bi_LSTM	0	0.96	0.97	0.96	4471	0.963
	1	0.97	0.95	0.96	4331	
CNN-BiLSTM	0	0.97	0.96	0.97	4471	0.966
	1	0.96	0.97	0.97	4331	

4. CONCUSION AND FUTURE WORKS

In this paper, we have presented three prominent models, namely LSTM, Bi-LSTM, and CNN-BiLSTM, which are instrumental in identifying fabricated news. We build our dataset for Vietnamedse Fake news detection. The structure, training methodologies, and performance evaluation of these algorithms have been explicated using both the Vietnamese dataset "VNDF-Vietnamese-fake-news-datasets" and the English dataset "True, Fake" (translated into Vietnamese). Empirical findings demonstrate that CNN-BiLSTM exhibits the highest efficacy among the algorithms in detecting fake news. These research outcomes hold potential for practical application in automated tools aimed at detecting fabricated news, thus aiding online communities in mitigating the adverse consequences associated with false information. Nevertheless, accurately and expeditiously discerning the veracity of news articles in the Vietnamese language entails several challenges pertaining to source verification, grammatical structure, and stylistic analysis. In order to solve this problem, we have proposed the next research direction which is natural language inference and knowledge graph.

REFERENCES

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 selection. *Journal of Economic Perspectives, 31*(2), 211-236.

Elhadad, M. K., Li, K. F., & Gebali, F. (2020). A novel approach for selecting hybrid features from online news textual metadata for fake news detection.*In Advances on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2019)* 14 (pp. 914-925). Springer International Publishing.

Kollias, D., & Zafeiriou, S. (2020). Exploiting multi-cnn features in CNN, RNN based dimensional emotion recognition on the omg in-the-wild dataset. *IEEE Transactions on Affective Computing, 12*(3), 595-606.

Kavianpour, P., & Kavianpour, M., Jahani, E., & Ramezani, A. (2021). A CNN-BILSTM model with attention mechanism for earthquake prediction. *arXiv 2021. arXiv preprint arXiv:2112.13444*.

Le, D. T., Vu, X. S., To, N. D., Nguyen, H. Q., Nguyen, T. T., Le, L., … & Nguyen, H. D. (2020). ReINTEL: A multimodal data challenge for responsible information identification on social network sites. *arXiv preprint arXiv:2012.08895.*

Nguyen, D. Q., & Nguyen, A. T. (2020). PhoBERT: Pre-trained language models for Vietnamese. *arXiv preprint arXiv: 2003.00744*.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter, 19*(1), 22–36.

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

Zhou, X., Zafarani, R. Network-based fake news detection: A pattern-driven approach. *ACM SIGKDD explorations newsletter, 21*(2), 48-60.

Zhou, X., Wu, J., & Zafarani, R. (2020). Safe: similarity-aware multi-modal fake news detection (2020). *Preprint. arXiv, 200304981*, 2

Quoc, D. T., & Dung, Q. N. (2012). An automatic text summarization soluion for Vietnamese. *National conference, Hanoi.*