

Article

Sentiment Analysis for Vietnamese—Based Hybrid Deep Learning Models

Cach N. Dang ¹, María N. Moreno-García ², Fernando De la Prieta ³, Kien V. Nguyen ^{1,4} and Vuong M. Ngo ⁵

¹ Science and Technology Application for Sustainable Development (STASD) Research Group, Ho Chi Minh City University of Transport (UTH), Ho Chi Minh 70000, Vietnam; cach@ut.edu.vn

² Data Mining (MIDA) Research Group, University of Salamanca, Salamanca 37007, Spain; mmg@usal.es

³ Biotechnology, Intelligent Systems and Educational Technology (BISITE) Research Group, University of Salamanca, Salamanca 37007, Spain; fer@usal.es

⁴ Ho Chi Minh University of Banking, Ho Chi Minh 70000, Vietnam; kiennv_httql@hub.edu.vn

⁵ Ho Chi Minh City Open University, Ho Chi Minh City, Vietnam; vuong.nm@ou.edu.vn

* Correspondence: cach@ut.edu.vn

Abstract: Sentiment analysis of public opinion expressed in social networks has been developed into various applications, especially in English. Hybrid approaches are potential models for reducing sentiment errors on increasingly complex training data. This paper aims to test some hybrid deep learning models' reliability in some domains' Vietnamese language. Our research questions are to determine whether it is possible to produce hybrid models that outperform the Vietnamese language. Hybrid deep sentiment-analysis learning models are built and tested on reviews and feedback of the Vietnamese language. The hybrid models outperformed the accuracy of Vietnamese sentiment analysis on Vietnamese datasets. It contributes to the growing body of research on Vietnamese NLP, providing insights and directions for future studies in this area.

Keywords: hybrid deep learning models; sentiment analysis; machine learning

1. Introduction

Sentiment analysis, commonly called opinion mining, is a branch of natural language processing (NLP) that focuses on locating and obtaining subjective data from text. Due to its many uses, such as sentiment categorization, opinion summarizing, customer feedback analysis, and social media monitoring, sentiment analysis has recently attracted a lot of attention. As a result, sentiment analysis has become an increasingly effective tool for businesses, researchers, and legislators to understand public opinion and customer preferences better as more individuals worldwide share their ideas and feelings on various internet platforms.

In recent years, deep learning approaches have been significant sentiment analysis tasks for numerous languages, including English [1], Spanish [2], Thai [3], Persian [4], Chinese [5], Arabic [6], and Hindi [7]. These techniques have been shown to outperform traditional machine learning methods by automatically learning hierarchical representations of text data, capturing features, and addressing the challenges posed by the inherent complexity of natural languages.

However, more research needs to be done on sentiment analysis in Vietnamese, an Austroasiatic language spoken by nearly 100 million people. Vietnamese presents difficulties for NLP tasks because of its distinctive language features, which include a complicated system of diacritical marks, tonal variances, and generally free word order. Vietnamese is also regarded as a low-resource language in NLP, with few annotated datasets and pre-trained models readily available.

This study proposes a hybrid deep learning model for Vietnamese sentiment analysis that incorporates Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and Support Vector Machines (SVM) to close this gap. The following are the work's main contributions:

1. We present a hybrid deep learning model that effectively captures Vietnamese text features by combining CNN, LSTM, and SVM.
2. We comprehensively compare our proposed model with several baseline models, including traditional machine learning techniques and standalone deep learning models, demonstrating the superiority of our approach on benchmark Vietnamese sentiment analysis datasets.
3. Our work contributes to the growing body of research on Vietnamese NLP, providing insights and directions for future studies in this area.

The rest of the paper is organized as follows: Section 2 discusses related work in Vietnamese sentiment analysis, highlighting the limitations of existing approaches. Section 3 describes the proposed hybrid deep learning model in detail, including the CNN layer, LSTM layer, and attention mechanism. Section 4 presents the experimental setup and results, comparing our model with baseline models on benchmark datasets. Finally, Section 5 concludes the paper and discusses future directions for research.

2. Related Work

Previous studies on Vietnamese sentiment analysis have primarily focused on traditional machine learning techniques, such as Naïve Bayes [8], SVM [9], and Decision Trees [10]. Unfortunately, these models rely on handcrafted features, such as bag-of-words (BoW), term frequency-inverse document frequency (TF-IDF), and n-grams, which limit their generalizability and performance. Some studies also proposed rule-based techniques for Vietnamese sentiment analysis [11] or presented a lexicon-based method for sentiment analysis of Facebook comments in the Vietnamese language [12]. The authors in [12] proposed a lexicon-based method for sentiment analysis by building a Vietnamese emotional dictionary (VED), including five sub-dictionaries: noun, verb, adjective, and adverb. In addition, the authors in [13], [14], and [15] used NLP, entity relationships, and named entity ontology to recognize customer opinions and detect opinion spam in reviews of Vietnam e-commerce websites.

With the success of deep learning in various NLP tasks, researchers started exploring deep learning techniques for Vietnamese sentiment analysis. Vu et al. applied LSTM to Vietnamese sentiment analysis [15]. They presented variants of LSTM for Sentiment Analysis on Vietnamese students' Feedback Corpus that LSTM could effectively outperform traditional machine learning models. Nguyen et al. [16] also applied LSTM to span detection for Vietnamese aspect-based sentiment analysis.

CNN and LSTM are also investigated for Vietnamese sentiment analysis. Vo et al. [17] provide a novel and efficient way for integrating the advantages of CNN and LSTM in Vietnamese text. However, their work did not incorporate an attention mechanism, which has been shown to improve performance in other languages.

Recent studies have started to explore transfer learning and pre-trained language models for Vietnamese sentiment analysis. Nguyen et al. [18] proposed BERT fine-tuning method for Sentiment Analysis of Vietnamese on reviews about food and restaurant on Foody and product reviews on various e-commerce sites. Similarly, Truong et al. [19] used a pre-trained PhoBERT model for Vietnamese sentiment analysis, demonstrating promising results.

Despite these developments, research on Vietnamese sentiment analysis primarily focuses on single deep learning architectures, and the possibility of combining different architectures has yet to be investigated. To fill this gap, we present a hybrid deep learning model that combines CNN, LSTM, and SVM to extract Vietnamese text features.

3. Proposed Model

The proposed hybrid deep learning model consists of three main components: a word embedding layer, a CNN layer, and an LSTM layer with an attention mechanism, as proposed in the research [20] and using the research [21], but these models are applied to English languages. In this

paper, we apply two models on Vietnamese language. We start by using a pre-trained BERT model to create the feature vector. Then, we used the CNN and LSTM models in the following stages: BERT -> CNN -> LSTM or BERT -> LSTM -> CNN. In the model's final stage, we use an SVM instead ReLU (rectified linear activation function) function because of its efficiency in word processing, especially in high dimensional contexts, such as natural language processing. The architecture of hybrid deep learning models is shown in the Figure 1. Moreover, the details of these model connections, the connection process, and the data-processing flow are indicated in Tables 1 and 2.

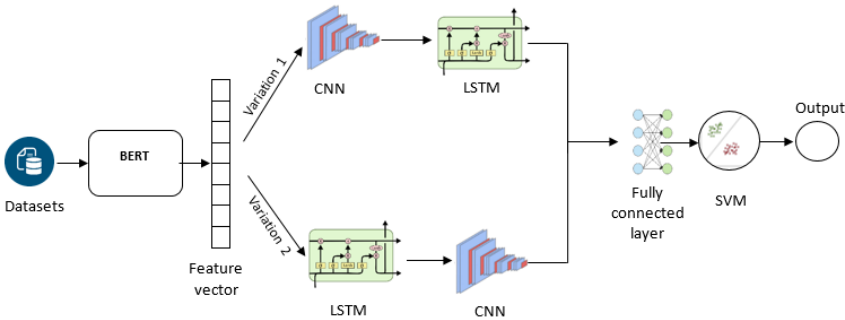


Figure. 1. Architecture of hybrid deep learning models for sentiment analysis.

Table 1. A Hybrid CNN-LSTM Model.

Layer (type)	Output Shape	Parameter #
BERT (base-uncased)	(None, 768, 1)	110,000,000
conv1d (Conv1D)	(None, 768, 512)	2,048
conv1d_1 (Conv1D)	(None, 768, 256)	393,472
conv1d_2 (Conv1D)	(None, 768, 128)	98,432
lstm_1 (LSTM)	(None, 500)	1,258,000
dense_1 (Dense)	(None, 50)	25,050
dense_2 (Dense)	(None, 2)	102
Total parameters: 111,777,104		
Trainable parameters: 111,777,104		
Non-trainable parameters: 0		

Table 2. A Hybrid LSTM-CNN Model.

Layer (type)	Output Shape	Parameter #
BERT (base-uncased)	(None, 768, 1)	110,000,000
lstm_1 (LSTM)	(None, 768, 500)	1,004,000
conv1d_1 (Conv1D)	(None, 768, 512)	768,512
conv1d_2 (Conv1D)	(None, 768, 256)	393,472
conv1d_3 (Conv1D)	(None, 768, 128)	98,432
flatten (Flatten)	(None, 98304)	0
dense_2 (Dense)	(None, 50)	4,915,250
dense_3 (Dense)	(None, 2)	102
Total parameters: 117,179,768		
Trainable parameters: 117,179,768		
Non-trainable parameters: 0		

3.1. Word Embedding Layer

The first step in our model is to convert the input text into a sequence of word embeddings. Word embeddings are continuous vector representations that capture semantic relationships between words. In this study, we use a pre-trained BERT model for Vietnamese. A pre-trained BERT

model is used in this study. After adjusting the parameters, the BERT model is used as a feature extractor to generate input data for the proposal of hybrid models. The Vietnamese datasets are fed into the BERT model to create the feature vectors. These embeddings are used as the input to the subsequent layers of the hybrid models.

3.2. CNN Layer

CNN is a deep neural network design that is usually made up of convolutional and pooling or subsampling layers that feed into a fully connected classification layer [22]. The CNN layer is designed to capture local features in the input text. It consists of multiple filters with varying sizes, which are applied to the word embeddings to capture different n-grams in the Vietnamese text. Each filter is followed by a rectified linear unit (ReLU) activation function to introduce non-linearity into the model. After applying the filters, we use max pooling to reduce the dimensionality of the feature maps and retain the most important features. The output of the CNN layer is a set of high-level feature representations that capture local semantic information in the input text.

A single convolutional (1D CNN) is employed in this study. The first layer of the hybrid model is the CNN, which receives the vector produced by word embedding. It has three convolution layers consisting of 512, 256, and 128 filters, respectively, with a kernel size = 3, which is used to reduce the output's complexity and prevent the overfitting of the data.

3.3. LSTM Layer with Attention Mechanism

The second layer of the hybrid model is the LSTM [23], which produces a matrix fed into the classifier. The LSTM layer captures the input text's long-range dependencies and global features. We employ LSTM cells, which process the input text in both forward and backward directions, allowing the model to effectively learn the context and relationships between words in a sentence. The output of the LSTM layer with attention mechanism is a weighted sum of the LSTM hidden states, which serves as a context vector that captures the global semantic information in the input text.

3.4. Fully Connected Layer and Output

The context vector generated by the LSTM layer with attention mechanism is passed through a fully connected (dense) layer with a ReLU activation function. This layer further processes the high-level features extracted by the CNN and LSTM layers, enabling the model to learn more complex relationships in the data.

Finally, the output of the fully connected layer with 128 nodes is fed into a SVM activation function to produce the final sentiment classification probabilities. The model is trained using categorical cross-entropy loss, and the sentiment label with the highest probability is selected as the final prediction.

4. Experimental Results

We implement our model using the Pytorch [24] and train it on Google Colab Pro with GPU Tesla P100-PCIE-16GB or GPU Tesla V100-SXM2-16GB [25]. The model is prepared using the Adam optimizer with a learning rate of 0.00005 and a batch size of 64. We use a dropout rate of 0.5 to prevent overfitting. The hyperparameters for the CNN and LSTM are selected through a grid search. Then, we compare our proposed model with several baseline single models, including SVM; deep neural networks (DNN) [26]; CNN; and LSTM. To evaluate the performance of the models, we use accuracy, precision, recall, F1-score, and AUC as the evaluation metrics.

4.1. Dataset and Preprocessing

We evaluate our proposed hybrid deep learning model on two benchmark Vietnamese sentiment analysis datasets:

1. USAL-UTH dataset collected by us: The USAL-UTH dataset consists of 10,000 Vietnamese customer comments collected from Shopee¹ e-commerce system. The reviews are labeled as positive or negative based on the rating, Negative with a 1-2 start, Positive with a five start.
2. UIT-VSFC dataset [27]: The UIT-VSFC dataset contains 16,000 Vietnamese students' feedback, with three classes: positive, neutral, and negative.

Figure 2 shows word cloud comments and feedback from the datasets described in Section 4.1. We perform tokenization, stop word removal, and diacritic normalization during preprocessing of the raw text. We then pad the sequences to a fixed length and convert them into the word embedding representations using PhoBERT [28] model for Vietnamese.



Figure 2. Word cloud of the combined dataset: (a) USL-UTH dataset in the first panel; (b) UIT-SVFC in the second panel.

4.2. Results and Discussion

The experimental results show that our proposed hybrid deep learning model achieves state-of-the-art performance on the USAL-UTH and UIT-VSFC datasets. Furthermore, the model outperforms the baseline models in terms of accuracy, precision, recall, F1-score, and AUC, as shown in Tables 3 and 4, and illustrated in Figures 3 and 4.

Table 3. The metrics comparison for the models with two classes (positive and negative) on UIT-SVFC dataset.

Models	Accuracy	Recall	Precision	F-Score	AUC
SVM	92.24	81.62	78.85	77.99	87.22
DNN	93.56	84.74	79.86	81.80	88.03
CNN	93.49	84.56	80.00	81.85	88.09
LSTM	93.42	82.93	80.74	81.69	88.49
LSTM-CNN	93.53	83.47	80.72	81.92	88.50
CNN-LSTM	93.52	83.82	80.61	82.01	88.43

¹ <https://shopee.vn/>

Table 4. The metrics comparison for the models with two classes (positive and negative) on USAL-UTH dataset.

Models	Accuracy	Recall	Precision	F-Score	AUC
SVM	93.33	94.54	92.57	93.52	93.35
DNN	93.35	93.38	93.90	93.63	93.36
CNN	93.56	94.06	93.58	93.81	93.56
LSTM	93.38	93.89	93.37	93.62	93.38
LSTM-CNN	93.21	94.17	92.73	93.42	93.21
CNN-LSTM	93.94	95.12	93.17	94.12	93.95

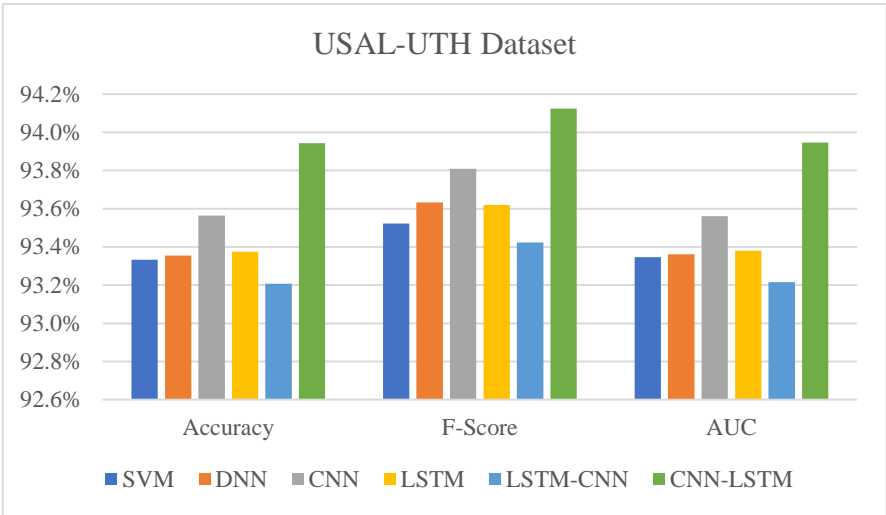


Figure 3. Accuracy, F_Score, and AUC values of the models with BERT for Shopee customer comments dataset.

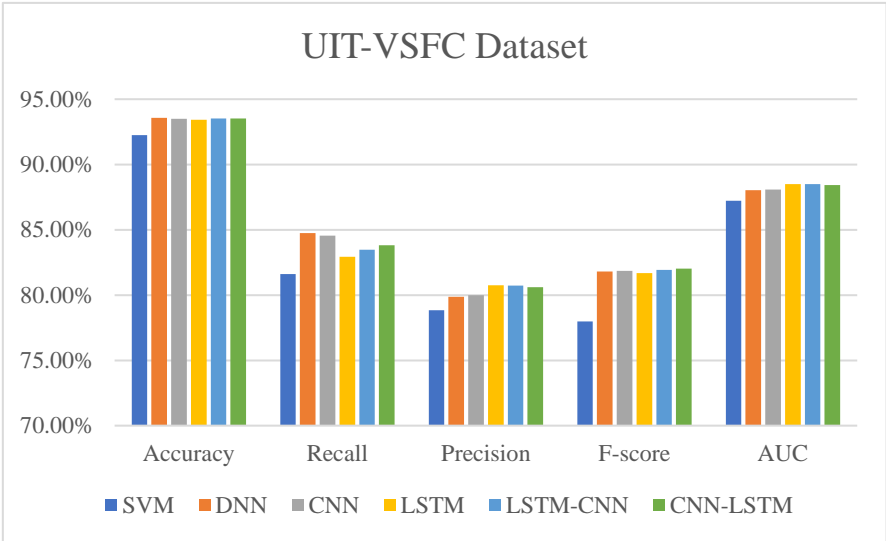


Figure 4. Accuracy, F_Score, and AUC values of the models with BERT for Students Feedback dataset.

For the USAL-UTH dataset, our model achieves an accuracy of 93.94%, an improvement of 0.59% over the standalone DNN model and 0.61% over the standalone SVM model. On the UIT-VSFC dataset, our model attains an accuracy of 93.53%, outperforming the standalone SVM model by 1.29%, respectively. In comparing sentiment analysis models applied on USAL-UTH and UIT-VSFC datasets, the combined CNN-LSTM are high value in all metrics (see Figure 5).

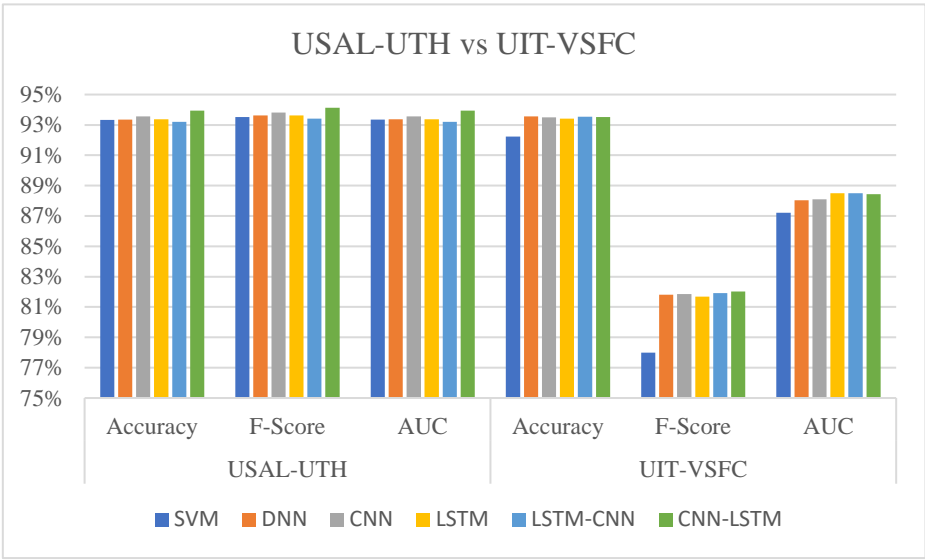


Figure 5. Comparing Accuracy, F_Score, and AUC values with BERT on USAL-UTH and UIT-VSFC datasets.

These results demonstrate the effectiveness of our proposed hybrid deep learning model in capturing features of Vietnamese text. The combination of CNN, LSTM, and SVM allows the model to learn complex relationships between words and their contexts, leading to superior performance in sentiment analysis tasks.

The performance gap between our model and traditional machine learning techniques is even larger, highlighting the advantages of deep learning approaches in sentiment analysis for Vietnamese.

5. Conclusions and Future Work

In this paper, we proposed a hybrid deep learning model for sentiment analysis in Vietnamese that combines CNN, LSTM, and SVM. Our model effectively captures features of Vietnamese text and outperforms on benchmark Vietnamese sentiment analysis datasets. This work contributes to the growing body of research on Vietnamese NLP and provides insights for future studies in this area.

Despite the promising results, there is stillroom for improvement and further exploration. Potential future work directions include exploring other deep learning architectures; Incorporating pre-trained language models; Leveraging transfer learning and domain adaptation; Expanding to other NLP tasks; Developing large-scale annotated datasets. By addressing these challenges and exploring these future directions, we believe our work can pave the way for more advanced and effective NLP models for Vietnamese and other low-resource languages.

References

1. Chen, Y.: Convolutional neural network for sentence classification. University of Waterloo (2015)
2. Paredes-Valverde, M.A., Colomo-Palacios, R., Salas-Zárate, M.d.P., Valencia-García, R.: Sentiment analysis in Spanish for improvement of products and services: A deep learning approach. Scientific Programming 2017, (2017)
3. Vateekul, P., Koomsubha, T.: A study of sentiment analysis using deep learning techniques on Thai Twitter data. In: 2016 13th International joint conference on computer science and software engineering (JCSSE), pp. 1-6. IEEE, (2016)
4. Roshanfekar, B., Khadivi, S., Rahmati, M.: Sentiment analysis using deep learning on Persian texts. In: 2017 Iranian conference on electrical engineering (ICEE), pp. 1503-1508. IEEE, (Year)
5. Zhang, X., Zhao, J., LeCun, Y.: Character-level convolutional networks for text classification. Advances in neural information processing systems 28, (2015)
6. Alomari, K.M., ElSherif, H.M., Shaalan, K.: Arabic tweets sentimental analysis using machine learning. In: Advances in Artificial Intelligence: From Theory to Practice: 30th International Conference on Industrial

- Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2017, Arras, France, June 27-30, 2017, Proceedings, Part I 30, pp. 602-610. Springer, (2017)
7. Yadav, V., Verma, P., Katiyar, V.: Long short term memory (LSTM) model for sentiment analysis in social data for e-commerce products reviews in Hindi languages. *International Journal of Information Technology* 15, 759-772 (2023)
 8. Le, C.-C., Prasad, P., Alsadoon, A., Pham, L., Elchouemi, A.: Text classification: Naive Bayes classifier with sentiment lexicon. *IAENG International journal of computer science* 46, 141-148 (2019)
 9. Le, H.P., Nguyen, T.M.H., Nguyen, P.T., Vu, X.L.: Building a Large Syntactically-Annotated Corpus of Vietnamese. In: *The Third Linguistic Annotation Workshop (The LAW III)*, pp. 6p., (2009)
 10. Bang, T.S., Haruechaiyasak, C., Sornlertlamvanich, V.: Vietnamese sentiment analysis based on term feature selection approach. In: *Proc. 10th International Conference on Knowledge Information and Creativity Support Systems (KICSS 2015)*, pp. 196-204. (2015)
 11. Kieu, B.T., Pham, S.B.: Sentiment analysis for Vietnamese. In: *2010 Second international conference on knowledge and systems engineering*, pp. 152-157. IEEE, (2010)
 12. Trinh, S., Nguyen, L., Vo, M., Do, P.: Lexicon-based sentiment analysis of Facebook comments in Vietnamese language. *Recent developments in intelligent information and database systems* 263-276 (2016)
 13. Nguyen, P., Le, L., Ngo, V., Nguyen, P.: Using Entity Relations for Opinion Mining of Vietnamese Comments. *arXiv preprint arXiv:1905.06647* (2019)
 14. Nguyen, L., Pham, N., Ngo, V.M.: Opinion Spam Recognition Method for Online Reviews using Ontological Features. *arXiv preprint arXiv:1807.11024* (2018)
 15. Nguyen, V.D., Van Nguyen, K., Nguyen, N.L.-T.: Variants of long short-term memory for sentiment analysis on Vietnamese students' feedback corpus. In: *2018 10th International conference on knowledge and systems engineering (KSE)*, pp. 306-311. IEEE, (2018)
 16. Nguyen, K.T.-T., Huynh, S.K., Phan, L.L., Pham, P.H., Nguyen, D.-V., Van Nguyen, K.: Span detection for aspect-based sentiment analysis in Vietnamese. *arXiv preprint arXiv:2110.07833* (2021)
 17. Vo, Q.-H., Nguyen, H.-T., Le, B., Nguyen, M.-L.: Multi-channel LSTM-CNN model for Vietnamese sentiment analysis. In: *2017 9th international conference on knowledge and systems engineering (KSE)*, pp. 24-29. IEEE, (2017)
 18. Nguyen, Q.T., Nguyen, T.L., Luong, N.H., Ngo, Q.H.: Fine-tuning bert for sentiment analysis of vietnamese reviews. In: *2020 7th NAFOSTED conference on information and computer science (NICS)*, pp. 302-307. IEEE, (2020)
 19. Truong, T.-L., Le, H.-L., Le-Dang, T.-P.: Sentiment Analysis Implementing BERT-based Pre-trained Language Model for Vietnamese. In: *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, pp. 362-367. IEEE, (2020)
 20. Dang, C.N., Moreno-García, M.N., De la Prieta, F.: Hybrid deep learning models for sentiment analysis. *Complexity* 2021, 1-16 (2021)
 21. Dang, C.N., Moreno-García, M.N., De la Prieta, F.: Using hybrid deep learning models of sentiment analysis and item genres in recommender systems for streaming services. *Electronics* 10, 2459 (2021)
 22. Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K.: Convolutional neural networks: an overview and application in radiology. *Insights into imaging* 9, 611-629 (2018)
 23. Hochreiter, S., Schmidhuber, J.: LSTM can solve hard long time lag problems. *Advances in neural information processing systems* 9, (1996)
 24. <https://pytorch.org/>
 25. <https://colab.research.google.com/notebooks/pro.ipynb>
 26. Aggarwal, C.C.: *Neural networks and deep learning*. Springer 10, 3 (2018)
 27. Van Nguyen, K., Nguyen, V.D., Nguyen, P.X., Truong, T.T., Nguyen, N.L.-T.: UIT-VSFC: Vietnamese students' feedback corpus for sentiment analysis. In: *2018 10th international conference on knowledge and systems engineering (KSE)*, pp. 19-24. IEEE, (2018)
 28. Nguyen, D.Q., Nguyen, A.T.: PhoBERT: Pre-trained language models for Vietnamese. *arXiv preprint arXiv:2003.00744* (2020)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.