Sentiment Analysis in Code-Mixed Vietnamese-English Sentence-level Hotel Reviews

Dang Van Thin, Duong Ngoc Hao, Ngan Luu-Thuy Nguyen
University of Information Technology, Ho Chi Minh city, Vietnam
Vietnam National University Ho Chi Minh city, Vietnam
{thindv,haodn,ngannlt}@uit.edu.vn

Abstract

In recent years, there has been an increasing amount of research on code-mixed Sentiment Analysis (SA) tasks due to the evolution of social media platforms in a multilingual society. This paper presents a comprehensive study on the Vietnamese-English code-mixed SA task, including (1) releasing two semi-annotated En-Vi code-mixed datasets; (2) investigating the performance of different machine learning, deep learning, and transformer-based approaches. The experimental results demonstrated that fine-tuning the multilingual sentence-transformer LaBSE (Feng et al., 2022) achieves better performance than the remaining approaches on two of our code-mixed SA datasets. Our work is the first tempt to solve the code-mixed Vietnamese-English SA problem to the best of our knowledge.

1 Introduction

The diversity of discussion platforms, such as forums, e-commerce, and social media, allows users to express their opinions and comments. This growth makes it challenging for individuals and organizations to understand users' aggregated thoughts (Ligthart et al., 2021). Therefore, the task of sentiment analysis has received a lot of attention from the NLP community (Liu and Zhang, 2012).

Code-mixing is the phenomenon of mixing the vocabulary and syntax of two or multiple languages in the sentence (Lal et al., 2019). Due to the rise of multilingual environments, there is an increase in code-mixed written text. Unlike monolingual sentences (e.g., English), code-mixing is very challenging for traditional NLP architectures because of grammatical constructions and spelling mistakes. Therefore, there has been a dramatic increase in code-mixed problems (Pratapa et al., 2018; Rani et al., 2020; Patwa et al., 2020; Chakravarthi et al., 2022).

On the other hand, recent multilingual NLP research has attracted the community's attention on word-level (Ruder et al., 2019) and sentence-level representations (Artetxe and Schwenk, 2019; Feng et al., 2022). Besides, the growth of multilingual Transformer-based language models (Devlin et al., 2019; Conneau et al., 2020) brought benefits to many downstream NLP tasks. Therefore, these representation methods can be effective for code-mixed tasks, especially low-resource languages. This paper presents a study on the code-mixed Vietnamese-English data for the SA task. The reason why we choose the Vietnamese-English language is that most Vietnamese people use English as a second language (Doan et al., 2018). Our main contributions can be summarized as follows:

- We release two code-mixed SA datasets for the Vietnamese and English languages for the hotel domain.
- We investigate the effectiveness of different machine learning and deep learning approach on the code-mixed sentiment analysis task.
- We perform experiments to confirm whether fine-tuning the SOTA pre-trained transformers benefit the code-mixed dataset. The experimental results demonstrated that fine-tuning the LaBSE model (Feng et al., 2022) achieves the best results on our datasets.

2 Related Work

In recent years, there has been an increasing interest in code-mixing or code-switching NLP tasks, including hate speech detection (Rani et al., 2020), Part-of-Speech tagging (Pratapa et al., 2018), Language Identification (Aguilar and Solorio, 2020). Moreover, researchers have shown an increased interest in code-mixing SA, however, most datasets are annotated for high-resource languages such as Spanish-English (Patwa et al., 2020), Hindi-English (Swamy et al., 2022; Hande et al., 2020; Patwa et al., 2020), Malayalam-English (Chakravarthi et al., 2020), Persian-English (Sabri et al., 2021), Dravidian-English (Chakravarthi et al., 2022).

Most recent studies focus on the machine learning approach to address the code-mixed sentiment analysis task. The authors (Hande et al., 2020; Patwa et al., 2020; Chakravarthi et al., 2022) investigated the performance of different machine learning and deep learning methods such as Support Vector Machine (SVM), Naive Bayes (NB), Convolution Neural Network (CNN). Moreover, there are some studies (Pratapa et al., 2018; Singh and Lefever, 2020) that utilized the effectiveness of cross-lingual word embedding approaches to perform code-mixing data. Their experimental results showed that incorporating the multilingual embedding increases the performance of baseline methods. With the development of multilingual language models such as mBERT(Devlin et al., 2019), XLM-R(Conneau et al., 2020), there were a few studies (Younas et al., 2020; Gupta et al., 2021) investigated the effectiveness of fine-tuning pre-trained language models for code-mixed SA datasets.

For Vietnamese language, the most recent works conducted on monolingual SA tasks by fine-tuning pre-trained language models (Nguyen et al., 2020; Truong et al., 2020). Our work is the first tempt to solve the code-mixed Vietnamese and English SA tasks to the best of our knowledge. In this paper, we introduce two Vietnamese-English code-mixed datasets and provide the performance of various benchmark approaches, including classical machine learning (ML) with handcraft features or multilingual sentence representations, deep learning with cross-lingual word embeddings and pre-trained language models.

3 Data Collection

The scarcity of code-mixed sentiment analysis datasets limits current study for low-resource languages. To tackle this research gap, we create two Vietnamese-English code-mixed datasets for the hotel reviews.

The development of our datasets are based on the available annotated SA dataset (Duyen et al., 2014). However, we found that this dataset still has some confusion between polarity classes and contains meaningless sentences. Therefore, we filtered and re-annotated the sentiment polarity label to ensure the dataset's quality. General, the term "code-mixed" refers to placing and mixing of words, phrases, and morphemes of two or more languages in the same sentence[1]. Besides, we notice that people often use English idioms or phrases to express ideas in Vietnamese text. Therefore, we create two datasets in two ways as follows: (1) Translating the noun, verb, and adjective to English in the review (named as WordDataset); (2) Translating the extracted keyword to the corresponding phrase (named as KeyDataset).

To build the WordDataset, we use the Vietnamese POS Tagging[2] tool to extract the Noun, Verb, and Adjective in the dataset to create the bilingual dictionary - where each word in this dictionary is translated to English. However, it is worth noticing that a word can have many different corresponding vocabularies in the target language; therefore, we expand this dictionary by adding its synonyms. For example, the word "khách sạn" can be translated as "hotel", "resort", or "hostel". Notice that this bilingual dictionary is corrected by a human with a language background in the hotel domain. Then, we randomize words in a set of extracted words (noun, verb, adjective) in each sentence to be replaced with a randomly corresponding translation word in the bilingual dictionary. This helps us create a dataset with different code-mixed sentences and is suitable for real applications because users can use different vocabulary in foreign languages.

In addition, instead of replacing essential words in the sentence, we also created another dataset based on translating the main keywords in the review. To extract the keyword, we use KeyBERT[3] based on the monolingual pretrained PhoBERT embedding (Nguyen and Tuan Nguyen, 2020). In order to distinguish to WordDataset, we extract keyphrases with the length ranging from 2 to 4 vocabularies. We also limit the number of top keyphrases in the long review to a value of 5. We build a keyphrase dictionary similar to the above

[1] https://en.wikipedia.org/wiki/Code-mixing
[2] https://github.com/trungtv/pyvi
[3] https://github.com/MaartenGr/KeyBERT

Table 1: Summary statistics for two datasets. Length is the average sentence length, Vocab is the size of the vocabulary.

	N.o sentences per class			Length	Vocab
	Positive	Negative	Neutral		
WordDataset	1981	780	543	14.38	47519
KeyDataset	1981	780	543	14.65	48397

procedure, where each value is translated to the target language and checked manually. Then, we randomly replace the extracted keywords in the review to create the new code-mixed dataset.

To ensure the quality of structural naturalness and lexical diversity in the code-mixed sentence, we conduct a revision process based on multilingual annotators to check and correct the dataset. The pseudocode to create word code-mixed datasets is illustrated in Algorithm 1. The detailed statistics of the two datasets are shown in Table 1.

Algorithm 1 Building Vietnamese-English word code-mixed SA Dataset.

Input: Vietnamese SA data: $S = \{s_n\}_{n=1}^{N}$; a Vi-En dictionaries: $dict = \{k; v\}_{m=1}^{M}$ where $v = \{v_1, v_2, ..., v_k\}$;
Output: Vi-En code-mixed dataset: $T = \{t_n\}_{n=1}^{N}$

```
for n ← 1...N do
    w_rep ← [];
    list_rep ← [];
    for m ← 1...M do
        if k_m in s_n then
            w_rep.insert(k_m);
            list_rep.insert(dict[k_m]) ;
        end
    end
    L ← Length(w_rep);
    for i ← 1...L do
        w ← random(list_rep[i]);
        s_n ← replace(w_rep[i], w);
    end
    t_n ← s_n;
end
```

4 Methods

This research aims to investigate the performance of different approaches for Vi-En code-mixed SA. Therefore, we conduct extensive experiments based on various methods, which we explain below.

Classical ML models + handcraft features: As a first baseline, we explore the performance of classical machine learning techniques, including SVM and Multilayer Perceptron. We extract the list of handcraft features (Duyen et al., 2014) and convert them to TF-IDF representation. The handcraft features are used as follows:

- **N-grams**: the bigrams of words is extracted as the features.
- **Important words**: We extract the main words in the review, including noun, verbs and adjectives.
- **POS Information**: All Part-of-Speech of words in the sentences.

Classical ML models + Sentence embedding: We also consider multilingual sentence embedding as the primary representation. In this case, we consider it as feature extraction and train them on ML classifiers. To the best of our knowledge, our study is the first attempt to explore the performance of multilingual embedding for the Vietnamese SA task. To extract the sentence representation, we investigate two newest cross-lingual sentence embeddings as below:

- **LASER**: LASER (Language-Agnostic Sentence Representations) is an encoder to generate pre-trained language representation in 93 languages, including very low-resource languages. It is able to map the sentences with the semantic closeness of different languages in a shared semantic vector space. The detail of LASER's architecture can be seen in the original work (Artetxe and Schwenk, 2019). This model achieved promising results for sentence-level NLP tasks, therefore, it is suitable to extract the code-mixed sentence representation.
- **LaBSE**: A related development is that of Language-agnostic BERT Sentence Embedding (LaBSE) (Feng et al., 2022). This model is based on a pre-trained BERT-like architecture, and dual encoder models create cross-lingual sentence embedding of 109 languages. The pre-trained model is also released to support downstream NLP tasks. In addition, this model can be used to represent parallel sentences through high-dimensional embeddings.

Deep learning approaches: Cross-lingual word embedding (Ruder et al., 2019) might be one of the interesting ways for the code-mixed problem. For the code-mixed SA tasks, there are some previous studies (Ma et al., 2020; Younas et al., 2020) which applied the deep learning models combined with various pre-trained cross-lingual embeddings such as MUSE (Lample et al., 2018), BPE(Heinzerling and Strube, 2018). As a result, we explore the effectiveness of two deep learning architectures (CNN (Kim, 2014) and LSTM) with different multilingual embeddings (MUSE and BPE) - which produced high performance in several cross-lingual NLP tasks.

- **MUSE**: MUSE is a toolkit that allows us to align the fastText word embeddings (Grave et al., 2018) in a common semantic space. We use pre-trained Vi-En mapping embedding to represent the words in a sentence and is updated during the training.

- **BPEMulti**: This multilingual subword embedding is trained on Wikipedia texts of 275 languages. This subword embedding is trained on a combination of data from multiple languages, and each subword is represented as 300 dimensions.

Transformer-based approaches: The recent development of transformer architectures has brought significant improvements to the NLP field. In the experiments, we consider three multilingual pre-trained language models, including mBERT (Devlin et al., 2019), XLM-R (Conneau et al., 2020), LaBSE (Feng et al., 2022). A short description of the models is given below:

- **mBERT**: Multilingual BERT is trained on 104 highest resource languages in Wikipedia data. Therefore, this model is able to produce cross-lingual representations that allow for fine-tuning the code-mixed sentence.

- **XLM-R**: This model utilized self-supervised training techniques and was trained on filtered Common Crawled data. Therefore, fine-tuning can significantly improve performance on a variety of cross-lingual benchmarks.

- **LaBSE**: LaBSE is introduced as a pre-trained multilingual language model which is trained based on two tasks: masked language modeling and translation language modeling for 109 languages. The experimental results demonstrated the performance of LaBSE in various NLP tasks(Feng et al., 2022). However, the effectiveness of fine-tuning this model has not been investigated for code-mixed tasks.

As the original work (Devlin et al., 2019), we fine-tune the pre-trained language model by putting the final hidden state **h** of [CLS] token as the representation of the code-mixed sentence. Then, a classifier with softmax activation is added to predict the probability of sentiment class **c**:

$$p(c|h) = softmax(Wh) \tag{1}$$

where **W** is the parameter matrix. The parameters of transformers and matrix **W** is updated during the training process.

5 Experiments

We use the stratified 5-folds cross-validation to report our experiments. The results are measured based on the micro-averaged and macro-averaged F1-score in all our experiments because of the imbalance in classes. Moreover, the weighted F1-score is used on the test set in previous studies (Patwa et al., 2020). We report the experimental results using different evaluation methods to compare the effectiveness of various approaches objectively.

As described in Section 4, we investigate and compare the performance of different models. For the classical ML model, we use the Linear SVM and the two layers MLP. The hyper-parameters of the two models are optimized using a grid search technique. We fix the architecture to 3 convolution layers with different kernel sizes (2,3,4), a dimensionality of 64 units, and ReLU activation for the CNN model. Then, we concatenate the output of global and max pooling features. For the LSTM model, we employ the bidirectional LSTM with 128 units and ReLU activation. To calculate the probability of polarities, we add two feed-forward layers with 300 dimensions, a ReLU activation for the first layer, and the number of classes dimensions and softmax activation for the second layer. We also apply dropout on the word embedding with a rate of 0.5 to prevent the overfitting of the two models. Two models are optimized using the Adam optimizer with a learning rate of 0.001, a batch size of 64, and a number of epochs of 100. For cross-lingual embeddings, we use the pre-trained MulBPE[4] with the size of 1 million vocabularies.

[4]https://bpemb.h-its.org/multi/

For the MUSE embeddings[5], we use the aligned fastText embeddings.

We used Huggingface's Trainer API (Wolf et al., 2020) to implement the transformers architectures, including mBERT[6], XLM-R[7] and LaBSE[8], and the hyperparameters were optimized using the search functionality offered by Trainer API. For the pre-processing component, we applied the same steps as a previous work (Thin et al., 2019).

5.1 Results and Analysis

Table 2 gives an overview of the results for classical ML models with handcraft features and sentence representations, while Table 3 shows the results of deep learning models combined with cross-lingual word embeddings and multilingual transformer-based language models.

As shown in Table 2, it can be noted that training models on LaBSE sentence representation consistently outperform other types of features in both datasets. It is obvious that handcraft features with TF-IDF representation achieved the lowest scores in terms of three F1 scores. One of the reasons for the poor performance of this approach is the sparsity of feature vectors and the diversity of vocabularies in both languages. Also, we observe that the performance of sentence representation is quite competitive with classical ML classifiers. Combining the sentence representation with ML classifier improves results than deep learning with cross-lingual embeddings. The reason might be because of the size of training data when the deep learning models often require more training data to achieve reasonably good performance. Comparing the results of SVM against MLP classifier shows that SVM yields better performance in types of features. Another interesting point is that ML models trained on LaBSE representation perform better than two remainder popular transformers models in terms of Weighted and Macro F1-score in both datasets. Our results demonstrated that LaBSE sentence embedding could produce an adequate representation for the code-mixed sentences. In addition, the results shown that the performance of models in both datasets is different; however, the difference is not significant.

We can also observe that fine-tuning the pre-trained SOTA multilingual sentence transformer

[5]https://github.com/facebookresearch/MUSE
[6]https://huggingface.co/bert-base-multilingual-cased
[7]https://huggingface.co/xlm-roberta-base
[8]https://huggingface.co/sentence-transformers/LaBSE

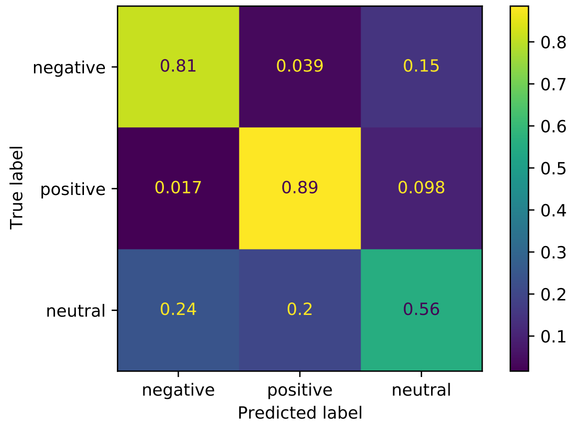

Figure 1: The confusion matrix of the LaBSE model on the WordDataset.

achieved the highest scores in all terms of F1-score in the two datasets. As seen in Table 3, the approach based on LaBSE outperformed more consistently than remainder methods in all evaluation metrics. Figure 1 and Figure 2 show the confusion matrix of LaBSE model on two datasets, respectively. We observe that the "neutral" label has the lowest score, while the two other classes achieve better results. The poor performance of the "neutral" class is because there are multiple sentences with opposite polarity; therefore, the label of these samples is annotated as "neutral" in original corpus(Duyen et al., 2014). For example, "khách sạn có vị trí đẹp nhưng nhân viên lễ tân giao tiếp còn kém, đồ ăn sáng thường." (*The hotel has a nice location, but the reception staff are not good at communicating, the breakfast is normal.*). The first phase in the sentence is expressed positively, while the remainder is the negative attitude of the user. That is why the overall sentiment polarity of these sentences is neural.

We also conducted an experiment to explore the performance of monolingual language models compared with the multilingual language model in the code-mixed data based on two scenarios on the WordDataset: (1) fine-tuning the latest pre-trained monolingual language models directly on code-mixed data (2) translating the code-mixed sentence to English and Vietnamese, then training translated data using monolingual language models. We choose the base version of PhoBERT (Nguyen and Tuan Nguyen, 2020), and RoBERTa (Liu et al., 2019) with the same above configuration for Vietnamese and English, respectively. We use Google API Translation to translate the code-mixed data

Table 2: The performance of classical machine learning models on two datasets.

Features	Model	KeyDataset			WordDataset		
		Weighted F1	Macro F1	Micro F1	Weighted F1	Macro F1	Micro F1
Handcraft + TFIDF	SVM	73.50	62.78	75.64	73.37	62.41	75.27
	MLP	73.60	63.67	74.36	73.39	63.25	73.88
Laser embedding	SVM	74.13	62.39	77.60	74.10	62.20	77.54
	MLP	74.83	65.52	75.27	75.41	66.13	75.48
LaBSE embedding	SVM	77.70	67.54	**79.72**	77.65	67.68	**79.54**
	MLP	**78.49**	**69.81**	78.72	**78.39**	**69.46**	79.21

Table 3: The performance of deep learning and transformers-based models on two datasets.

Approach	Model	KeyDataset			WordDataset		
		Weighted F1	Macro F1	Micro F1	Weighted F1	Macro F1	Micro F1
Deep learning	CNN + MUSE	75.61	66.47	76.06	74.79	64.86	75.30
	LSTM + MUSE	73.62	63.04	74.43	74.03	63.71	74.15
	CNN + MultiBPE	75.46	65.85	76.24	74.60	64.57	75.24
	LSTM + MultiBPE	73.58	64.11	73.76	73.25	63.24	73.46
Transformers model	mBERT	75.75	63.86	78.66	76.32	65.34	78.42
	XLM-R	76.23	63.59	79.60	76.26	63.98	79.75
	LaBSE	**81.87**	**73.38**	**82.90**	**82.24**	**74.04**	**83.05**

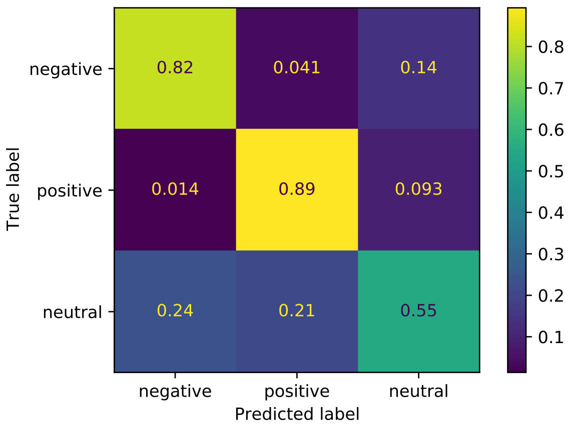

Figure 2: The confusion matrix of the LaBSE model on the KeyDataset.

to specific languages (English and Vietnamese) for the second scenario. Figure 3 shows the Weighted F1-score of two scenarios and the LaBSE's performance. It is obvious that fine-tuning directly pre-trained multilingual models gain better results than monolingual models in two scenarios for the code-mixed data. These experiments show that multilingual models are able to achieve competitive results for code-mixed tasks.

6 Conclusion

This paper presents a comprehensive Vietnamese and English code-mixed Sentiment Analysis for the hotel domain. Firstly, we introduced two code-mixed Vi-En datasets created based on the semi-

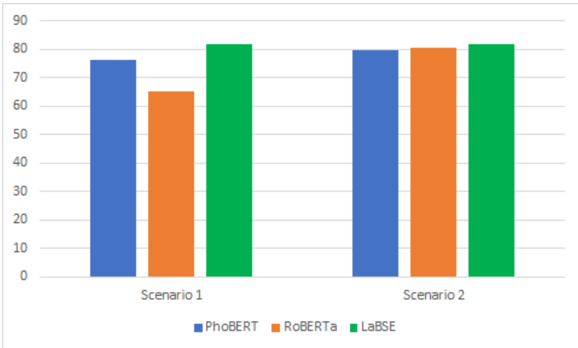

Figure 3: The performance of monolingual models compared with the LaBSE model in two scenarios.

approach. Secondly, we investigated the different methods on two datasets, including the classical ML approach combined with handcraft features or sentence representations, deep learning architectures with cross-lingual word embeddings, and SOTA multilingual language models. It is surprising that fine-tuning the pre-trained LaBSE (Feng et al., 2022) achieved the highest performance. We release two datasets and our code for the research community to facilitate future work on the code-mixed Vi-En SA task.

Acknowledgements

This research was supported by The VNUHCM-University of Information Technology's Scientific Research Support Fund.

References

Gustavo Aguilar and Thamar Solorio. 2020. From English to code-switching: Transfer learning with strong morphological clues. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8033–8044, Online. Association for Computational Linguistics.

Mikel Artetxe and Holger Schwenk. 2019. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7:597–610.

Bharathi Raja Chakravarthi, Navya Jose, Shardul Suryawanshi, Elizabeth Sherly, and John Philip McCrae. 2020. A sentiment analysis dataset for code-mixed Malayalam-English. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 177–184, Marseille, France. European Language Resources association.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Vigneshwaran Muralidaran, Navya Jose, Shardul Suryawanshi, Elizabeth Sherly, and John P McCrae. 2022. Dravidiancodemix: Sentiment analysis and offensive language identification dataset for dravidian languages in code-mixed text. *Language Resources and Evaluation*, 56(3):765–806.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Ngoc Doan, Toan Pham, Min Pham, and Kham Tran. 2018. English as an international language in viet nam: History and development. *Asian Englishes*, 20(2):106–121.

Nguyen Thi Duyen, Ngo Xuan Bach, and Tu Minh Phuong. 2014. An empirical study on sentiment analysis for vietnamese. In *Proceedings of ATC*, pages 309–314.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 878–891, Dublin, Ireland. Association for Computational Linguistics.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning word vectors for 157 languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Akshat Gupta, Sai Krishna Rallabandi, and Alan W Black. 2021. Task-specific pre-training and cross lingual transfer for sentiment analysis in Dravidian code-switched languages. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 73–79, Kyiv. Association for Computational Linguistics.

Adeep Hande, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2020. KanCMD: Kannada CodeMixed dataset for sentiment analysis and offensive language detection. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 54–63, Barcelona, Spain (Online). Association for Computational Linguistics.

Benjamin Heinzerling and Michael Strube. 2018. BPEmb: Tokenization-free pre-trained subword embeddings in 275 languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

Yash Kumar Lal, Vaibhav Kumar, Mrinal Dhar, Manish Shrivastava, and Philipp Koehn. 2019. De-mixing sentiment from code-mixed text. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 371–377, Florence, Italy. Association for Computational Linguistics.

Guillaume Lample, Alexis Conneau, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018. Word translation without parallel data. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.

Alexander Ligthart, Cagatay Catal, and Bedir Tekinerdogan. 2021. Systematic reviews in sentiment analysis: a tertiary study. *Artificial Intelligence Review*, 54(7):4997–5053.

Bing Liu and Lei Zhang. 2012. *A Survey of Opinion Mining and Sentiment Analysis*, pages 415–463. Springer US, Boston, MA.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Yili Ma, Liang Zhao, and Jie Hao. 2020. XLP at SemEval-2020 task 9: Cross-lingual models with focal loss for sentiment analysis of code-mixing language. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 975–980, Barcelona (online). International Committee for Computational Linguistics.

Dat Quoc Nguyen and Anh Tuan Nguyen. 2020. PhoBERT: Pre-trained language models for Vietnamese. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1037–1042, Online. Association for Computational Linguistics.

Quoc Thai Nguyen, Thoai Linh Nguyen, Ngoc Hoang Luong, and Quoc Hung Ngo. 2020. Fine-tuning bert for sentiment analysis of vietnamese reviews. In *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, pages 302–307.

Parth Patwa, Gustavo Aguilar, Sudipta Kar, Suraj Pandey, Srinivas PYKL, Björn Gambäck, Tanmoy Chakraborty, Thamar Solorio, and Amitava Das. 2020. SemEval-2020 task 9: Overview of sentiment analysis of code-mixed tweets. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 774–790, Barcelona (online). International Committee for Computational Linguistics.

Adithya Pratapa, Monojit Choudhury, and Sunayana Sitaram. 2018. Word embeddings for code-mixed language processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3067–3072, Brussels, Belgium. Association for Computational Linguistics.

Priya Rani, Shardul Suryawanshi, Koustava Goswami, Bharathi Raja Chakravarthi, Theodorus Fransen, and John Philip McCrae. 2020. A comparative study of different state-of-the-art hate speech detection methods in Hindi-English code-mixed data. In *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, pages 42–48, Marseille, France. European Language Resources Association (ELRA).

Sebastian Ruder, Ivan Vulić, and Anders Søgaard. 2019. A survey of cross-lingual word embedding models. *J. Artif. Int. Res.*, 65(1):569–630.

Nazanin Sabri, Ali Edalat, and Behnam Bahrak. 2021. Sentiment analysis of persian-english code-mixed texts. In *2021 26th International Computer Conference, Computer Society of Iran (CSICC)*, pages 1–4.

Pranaydeep Singh and Els Lefever. 2020. Sentiment analysis for Hinglish code-mixed tweets by means of cross-lingual word embeddings. In *Proceedings of the The 4th Workshop on Computational Approaches to Code Switching*, pages 45–51, Marseille, France. European Language Resources Association.

Sowmya Swamy, Jyoti Kundale, and Dipti Jadhav. 2022. Sentiment analysis of multilingual mixed-code, twitter data using machine learning approach. In *Proceedings of ICICC*, pages 683–697.

Dang Van Thin, Duc-Vu Nguyen, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen, and Anh Hoang-Tu Nguyen. 2019. Multi-task learning for aspect and polarity recognition on vietnamese datasets. In *Proceedings of PACLING*, pages 169–180.

Trong-Loc Truong, Hanh-Linh Le, and Thien-Phuc Le-Dang. 2020. Sentiment analysis implementing bert-based pre-trained language model for vietnamese. In *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, pages 362–367.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Aqsa Younas, Raheela Nasim, Saqib Ali, Guojun Wang, and Fang Qi. 2020. Sentiment analysis of code-mixed roman urdu-english social media text using deep learning approaches. In *2020 IEEE 23rd International Conference on Computational Science and Engineering (CSE)*, pages 66–71.