Unified Hope Speech Detection Across Languages: A LaBSE-Based Approach for IberLEF 2025

Tolulope Olalekan Abiola[1,*,†], Olumide Ebenezer Ojo[1,†], Olaronke Oluwayemisi Adebanji[1,†], Oluwatobi Joseph Abiola[2,†], Temitope Dasola Ogunleye[3,†] and Sidorov Grigori[1,†]

[1]*Centro de Investigación en Computación, Instituto Politécnico Nacional, CDMX, Mexico.*
[2]*Federal University Oye-Ekiti, Ekiti, Nigeria*
[3]*Ladoke Akintola University of Technology, Ogbomoso, Nigeria*

Abstract

This paper proposes a multilingual approach to hope speech detection through the Language-agnostic BERT Sentence Embedding (LaBSE) model. In compliance with the need for language-agnostic and scalable classification, we evaluate our approach on the English and Spanish versions of the PolyHope IberLEF 2025 dataset. Our approach ranks competitively as the second best for Spanish and sixth best for English among participating systems. Experimental results show that LaBSE outperforms traditional models with a unified framework for cross-lingual hope speech classification with task-specific fine-tuning. The research emphasises the potential of multilingual transformers for inclusive and affective content analysis in diverse linguistic settings.

Keywords

Hope Speech, Multilingual NLP, LaBSE, Transformer Models, Sentiment Analysis, IberLEF 2025

1. Introduction

In the internet era, social media platforms such as Twitter have become central to how humans communicate, express emotions, and create communities across the globe. Social media websites are not merely sites of information sharing but also emotional spaces where individuals share experiences of joy, grief, resilience, and hope. While a great deal of the natural language processing (NLP) research has been focused on detecting noxious or toxic content—hate speech, abuse, and misinformation, for instance—there is an increasing recognition that there should also be detection and promotion of positive and constructive communication. Hope speech is one of these positive communications.

Hope speech is described as messages that convey encouragement, optimism, support, and an anticipation of good things. It plays a key role in building mental wealth, cementing social bonds, and combating the toxicity that seems to prevail in online spaces. Hope speech is particularly vital in times of crisis, e.g., pandemics, political unrest, or personal loss, since it has the potential to strengthen communities and provide psychological comfort. Despite its utility, automatic hope speech detection is a difficult task with some unique challenges. Hope may be indirect or figurative; it may be presented in figures of speech, conditioned by cultural context, or reliant on subtle semantic cues.

To this end, shared tasks on hope speech detection have been suggested to encourage the development of computational systems that identify and amplify hopeful communication. This paper focuses on Subtask 1: Binary Hope Speech Detection, which aims to classify social media texts into two categories:

- **Hope:** Tweets containing an expression of hope, expectation, or desire.
- **Not Hope:** Tweets that do not contain hope, expectation, or desire.

IberLEF 2025, September 2025, Zaragoza, Spain

*Corresponding author.
†These authors contributed equally.
abiolato92@gmail.com (T. O. Abiola); olumideoea@gmail.com (O. E. Ojo); oluwatobiabiola01@gmail.com (O. J. Abiola); ogunleyetemitopedasola@gmail.com (T. D. Ogunleye); sidorov@cic.ipn.mx (S. Grigori)
0009-0005-2817-6264 (T. O. Abiola)

Our research aims at two languages—English (Subtask 1.a) and Spanish (Subtask 1.b)—to investigate the appearance of hope speech in various linguistic and cultural environments. The binary classification task serves as the building block, providing the basis for more complex detection of positive discourse in multilingual and multicultural environments.

An essential aspect of this task is the identification of not just overt expressions of hope but also implicit or metaphorical ones. Social media users prefer to express hope in indirect or figurative terms, which makes it harder for computers to process. This necessitates models to be sensitive to linguistic nuance and capable of reading between the lines of sentiment beyond literal meaning.

To tackle these difficulties, researchers have begun exploring multilingual and language-agnostic approaches to hope speech detection, drawing on the potential of modern NLP models to perform well across a range of language inputs. Our work belongs to this nascent trend by undertaking a multilingual binary classification task with a focus on exploring how hopeful sentiment is conveyed and can be computationally detected in both English and Spanish tweets.

The remaining portion of this paper outlines our methodology, data preprocessing pipeline, modelling strategies, and evaluation procedure. Last but not least, this research continues efforts in developing emotionally intelligent systems that prioritize empathy, inclusion, and psychological well-being in virtual environments.

2. Literature Review

Text classification remains a central task in Natural Language Processing (NLP), evolving from traditional machine learning (ML) techniques to sophisticated deep learning (DL) and transformer-based architectures. Foundational studies employed models such as Support Vector Machines, Decision Trees, and Naïve Bayes for tasks ranging from topic classification to sentiment analysis [1, 2, 3]. These methods laid the groundwork for more context-aware classification systems. Subsequently, neural approaches, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), demonstrated enhanced capabilities in capturing syntactic and semantic patterns within text [4, 5, 6, 7, 8]. The rise of transformer models such as BERT and its multilingual variants brought further performance gains, particularly in multilingual and low-resource contexts [9, 10, 11, 12, 13, 14].

Within this broader context of text classification, hope speech detection has emerged as a distinct research focus. Positioned as a counterbalance to the well-established domain of hate speech detection, hope speech centers on identifying content that is optimistic, inclusive, and supportive. Its significance lies in promoting positive discourse on social media platforms and supporting mental health and social cohesion. The task has seen increasing scholarly and shared-task attention, particularly within the framework of the HOPE track at IberLEF [15, 16, 17, 18, 19, 20, 21].

Initial efforts in the field focused on binary classification of hope speech using traditional methods. For example, Yigezu et al. [22] applied Support Vector Machines on English and Spanish datasets, reporting modest F1-scores of 0.489 and 0.481. As transformer-based models gained prominence, researchers began leveraging their power for multilingual applications. Divakaran et al. [23, 24] developed models such as GUIDE4Hope, incorporating BERT and TF-IDF-based logistic regression to achieve macro F1-scores up to 0.82 for English and Spanish binary classification. They also explored multiclass classification, reaching 0.64 macro F1-score.

A major advancement in multilingual modelling was achieved by Ahmad et al. [25, 26], who introduced the Posi-Vox-2024 dataset in English, Urdu, and Arabic. Their transfer learning approach using BERT demonstrated F1-scores of 0.78 in binary classification and improved accuracy over logistic regression baselines, particularly for low-resource languages. Further expanding on multilingual detection, Sharma et al. [27] proposed an ensemble model combining LSTM, mBERT, and XLM-RoBERTa for English, Kannada, Malayalam, and Tamil, achieving high weighted F1-scores of 0.93, 0.74, 0.82, and 0.60, respectively.

Some studies have also focused on novel or low-resource language datasets. For instance, Nath et al. [28] addressed the scarcity of Bengali-language resources by introducing BongHope, a binary-labeled dataset for positive discourse detection in Bengali social media. Arif et al. [29] brought a psycholinguistic perspective to the task, applying lexicon-based tools such as LIWC, NRC-emotion-lexicon, and VADER to identify cognitive and emotional indicators of hope. Their use of LightGBM and CatBoost achieved competitive results, emphasising the value of lightweight ML methods.

A particularly notable contribution to the multilingual landscape came from Balouchzahi et al. [30], who developed an Urdu-language dataset with a dual focus on hope and hopelessness. Their semi-supervised annotation strategy—combining large language models (LLMs) and human review—enabled nuanced classification into multiple emotional categories. Their best-performing transformer-based models achieved macro F1-scores of 0.4801 in multiclass tasks. Similarly, Armenta-Segura and Sidorov [31] developed custom multilingual BERT models trained on Spanish and English hope speech content from the HopeEDI and PolyHope datasets, incorporating multilingual sentiment embeddings and achieving top rankings in the HOPE@IberLEF 2024 shared task.

The HOPE shared tasks themselves have been pivotal in catalyzing advancements in this field. García-Baena et al. [20] curated the SpanishHopeEDI dataset focusing on LGBT-related content and evaluated a range of baselines, while García et al. [19] provided a comprehensive overview of the IberLEF 2024 task, noting participation from 19 teams and reporting macro F1-scores exceeding 0.78 in multiclass configurations.

Alternative modeling approaches have also been proposed. Arunadevi et al. [32] used logistic regression with TF-IDF and count vectorizer features for Spanish, achieving a macro-F1 score of 0.4161. Eyob et al. [33] compared transformer models with traditional algorithms such as SVM and Random Forest, showing that BERT models substantially outperform conventional techniques, reaching macro F1-scores of 0.85. Ullah et al.

Additional work by Junaida and Ajees [34] explored hope speech classification in Tamil and Malayalam, where low-resource constraints are particularly pronounced. Their experiments demonstrated the effectiveness of RNNs and contextual embeddings, with Iyer et al. achieving F1-scores of 0.93 for English, 0.58 for Tamil, and 0.84 for Malayalam using a context-aware deep learning model.

Butt et al. [35] present a comprehensive study on multi-class hope speech detection in both Spanish and English, exploring optimism, expectation, and sarcasm classification. The authors evaluate multiple transformer-based models using fine-tuning and prompt-learning strategies. For the Spanish dataset, ALBERT achieved the highest macro F1-score (0.8425), slightly outperforming RoBERTa (0.8401), while DistilBERT yielded the lowest performance (0.7697). In the English binary hope speech detection task, prompt-learning methods using LLaMA3 and GPT-4 were assessed under both few-shot learning (FSL) and zero-shot learning (ZSL) settings. GPT-4 with FSL achieved the best overall performance with a macro F1-score of 0.7823, outperforming LLaMA3 FSL (0.7401) and both models under ZSL. These findings demonstrate the effectiveness of large language models in hope speech detection, particularly under few-shot learning settings.

Despite these contributions, most existing research is limited to language-specific or bilingual classification tasks, often requiring separate pipelines or model retraining per language. Only a few have explored unified multilingual models capable of generalising across structurally diverse languages. Addressing this gap, our study proposes the use of Language-agnostic BERT Sentence Embedding (LaBSE) to create a single semantic embedding space for binary hope speech classification across four typologically distinct languages: English, and Spanish. This approach builds on the successes of multilingual transformers while aiming to reduce the architectural fragmentation and annotation overhead common in prior work.

3. Methodology

3.1. Datasets

The datasets used in this study are provided by the PolyHope at IberLEF 2025 [36]: *Optimism, Expectation, or Sarcasm?* shared task organisers. These datasets consist of text data labelled as *Hope* or *Not Hope*, and were not publicly available before this task. The primary goal is to classify the text into binary categories based on whether it exhibits hopeful or non-hopeful sentiment.

Two datasets were used in this study: one for English and another for Spanish. The English dataset contains 5,307 instances of hopeful text and 5,927 instances of non-hopeful text. The Spanish dataset, which is more balanced, includes 2,426 instances of hopeful text and 2,807 instances of non-hopeful text. These datasets are split into training, validation, and test sets.

The class distribution for both the training and validation sets, along with average word counts for each class, is presented in Table 1. As shown, the English dataset has a higher number of instances for the *Not Hope* class in comparison to the *Hope* class. Similarly, the Spanish dataset also shows a slightly higher number of non-hopeful texts compared to hopeful ones.

In addition, the distribution of word counts across the classes for both the English and Spanish datasets is shown in Figure 1. As seen in the figure, the average word count is slightly higher for the *Not Hope* class in both languages. Specifically, the English dataset has an average word count of 36.02 for *Not Hope* and 31.17 for *Hope*. For the Spanish dataset, the average word count is 35.42 for *Not Hope* and 31.56 for *Hope*.

Language	Class	Train Count	Dev Count	Avg Word Count (Train)
English	Not Hope	5,927	2,807	36.02
	Hope	5,307	2,426	31.17
Spanish	Not Hope	2,807	1,003	35.42
	Hope	2,426	899	31.56

Table 1
Class Distribution and Average Word Count for English and Spanish Datasets

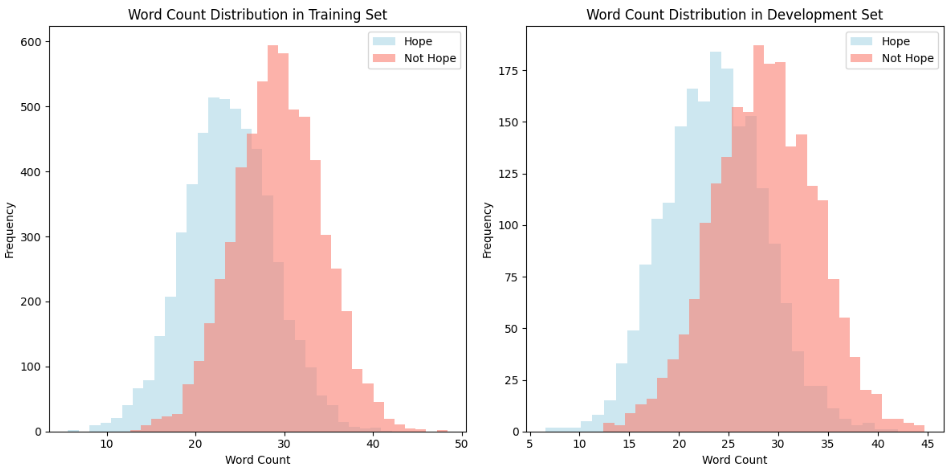

Figure 1: Distribution of Average Word Count for *Hope* and *Not Hope* Classes in English and Spanish Datasets

3.2. Preprocessing

Prior to training the models, extensive text preprocessing was performed. First, all text data was converted to lowercase to ensure uniformity across different text formats. Then, non-alphanumeric characters, including punctuation and special symbols, were removed using regular expressions. Additionally, URLs and links and other non-word characters like emojis were also eliminated to reduce noise in the data, as they were not deemed relevant for the sentiment classification task. These preprocessing steps were applied to both the English and Spanish datasets.

3.3. Model Architecture

We explored two primary models for classifying hope speech: a traditional machine learning approach using Random Forest classifiers, and a transformer-based architecture utilising LaBSE (Language-agnostic BERT Sentence Embedding). The Random Forest model, based on TF-IDF features extracted from the preprocessed text, served as a baseline model. On the other hand, LaBSE, which is a multilingual pre-trained model capable of handling multiple languages, was fine-tuned on the task-specific datasets.

3.3.1. Random Forest Classifier

For the Random Forest model, we did a TF-IDF vectorizer to transform the text data into feature vectors and subsequently trained a Random Forest classifier. The Random Forest model was trained on 100 trees with maximum tree depth and minimum sample split size of 2 to allow the model to learn the subtle patterns from the text data. Hyperparameters such as the number of estimators and tree depth were tuned via experimentation.

3.3.2. LaBSE Model

LaBSE is transformer-based multilingually pre-trained model in various languages. It was also fine-tuned specifically for the binary classification task of hope speech detection. LaBSE was also pre-trained using the tokenized text of Spanish and English datasets. The batch size was used as 32 and the learning rate was 2×10^{-5}. Training was carried out for 5 epochs that runs for 5hours 35mins for the training period. This model was experimentally tested for its ability to classify hopeful vs. non-hopeful text in both English and Spanish languages.

3.4. Evaluation and Prediction

The models were evaluated with standard classification measures like accuracy, precision, recall, and F1-score. Confusion matrices were also built to be able to compare the performance of both models on the validation sets. These measures were calculated for the English as well as Spanish datasets and revealed a clear contrast of how good each model performed in identifying hope speech.

For the final predictions, the best-performing models were run on the test data. The prediction results were submitted to the organisers for evaluation, using the test set to generate the official metrics as per the organisers' scoring system.

4. Results

4.1. Development Set Results

During the development phase, both the Random Forest (RF) and LaBSE models were evaluated using the validation datasets for English and Spanish. The performance metrics, including precision, recall, and accuracy, are summarised in Table 2. As shown, LaBSE consistently outperformed the RF model across both languages.

For the English validation set, the RF model achieved an accuracy of 78%, whereas LaBSE attained a higher accuracy of 85%. Figure 2 illustrates the corresponding confusion matrices, where LaBSE demonstrates a more balanced classification performance. Similarly, on the Spanish validation set, RF reached an accuracy of 78%, while LaBSE improved upon this with an accuracy of 83%, as visualized in Figure 3.

These results highlight the robustness and superior generalisation capability of the LaBSE model across both monolingual datasets.

Language	Model	Precision	Recall	Accuracy
English	RF	0.78	0.78	0.78
English	LaBSE	0.85	0.85	0.85
Spanish	RF	0.78	0.78	0.78
Spanish	LaBSE	0.83	0.83	0.83

Table 2
Development Set Evaluation Metrics for English and Spanish

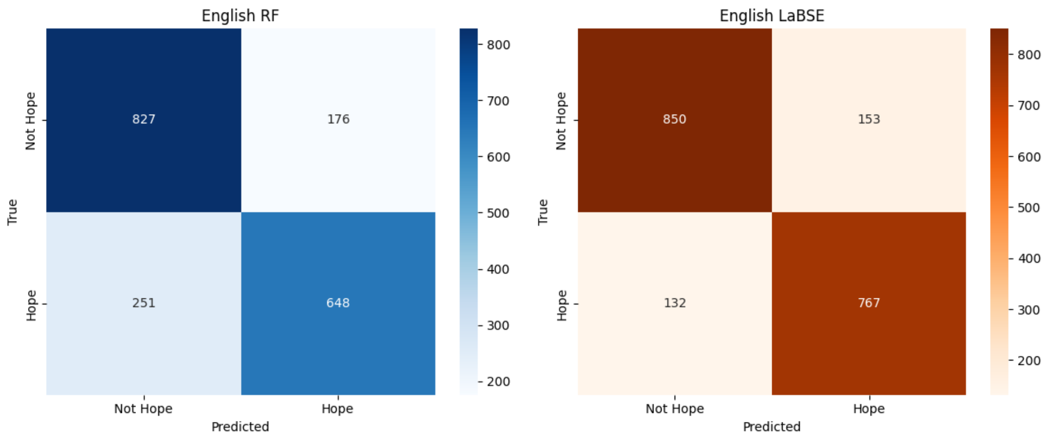

Figure 2: Confusion Matrices for English Validation Set — RF and LaBSE

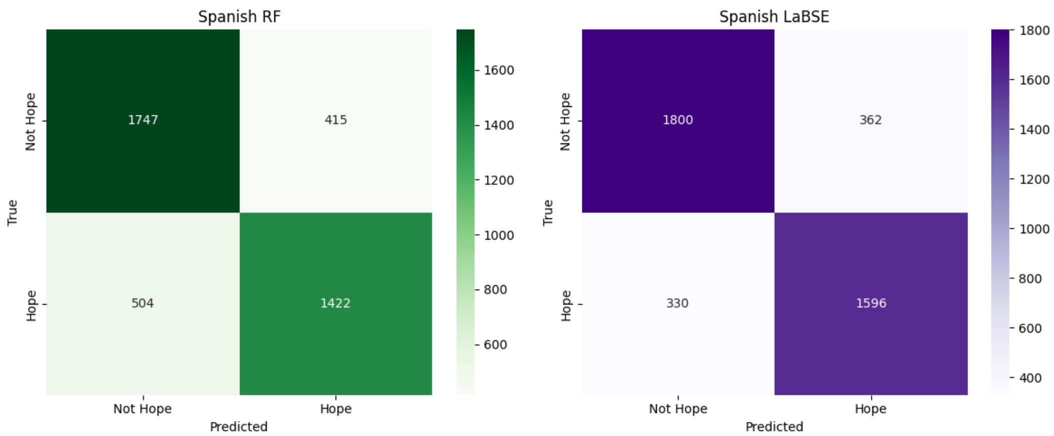

Figure 3: Confusion Matrices for Spanish Validation Set RF and LaBSE

4.2. Test Set Results

The final results submitted to the PolyHope IberLEF 2025 evaluation platform are summarised in Table 3. These results represent the performance of the Random Forest (RF), Logistic Regression (LR), and LaBSE models on the official test sets for English and Spanish. The LaBSE model consistently outperformed the classical baselines across all metrics—weighted F1, macro F1, accuracy, and precision.

Figure 4 provides a visual summary of the performance trends across key evaluation metrics for the English and Spanish test sets. LaBSE consistently leads in all measured categories.

Figure 5 further illustrates a side-by-side comparison of performance by language and model, highlighting the multilingual generalisation advantage of LaBSE.

Language	Model	Weighted F1	Macro F1	Accuracy	Precision
English	RF	0.795	0.794	0.796	0.796
English	LaBSE	**0.867**	**0.866**	**0.867**	**0.867**
Spanish	RF	0.780	0.779	0.781	0.781
Spanish	LaBSE	**0.846**	**0.846**	**0.846**	**0.855**

Table 3
Performance on the test set submitted to the PolyHope IberLEF 2025 platform. LaBSE outperforms classical models across all key metrics.

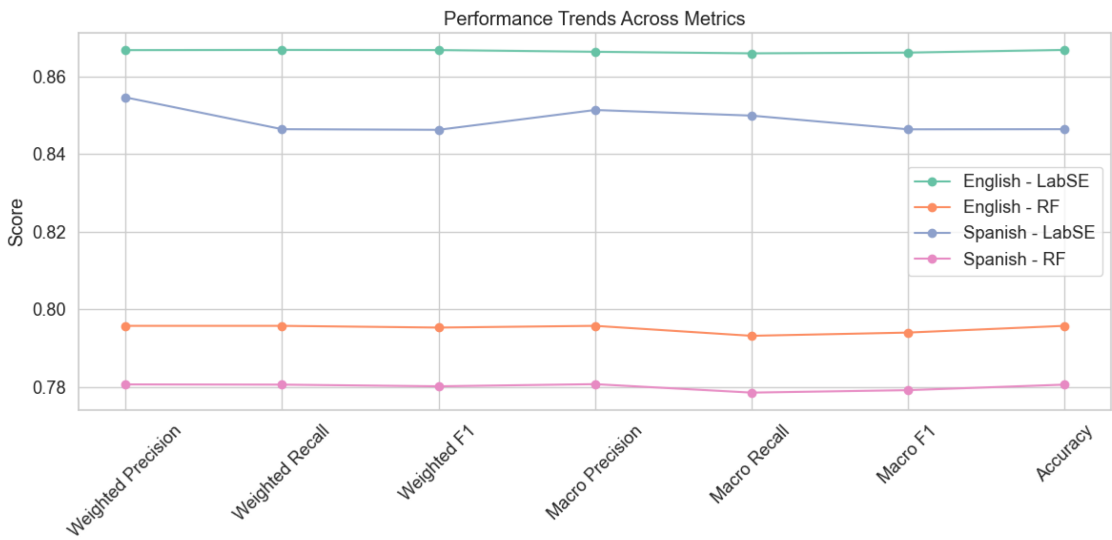

Figure 4: Performance trend across evaluation metrics for English and Spanish datasets. LaBSE achieves consistently higher values in all categories.

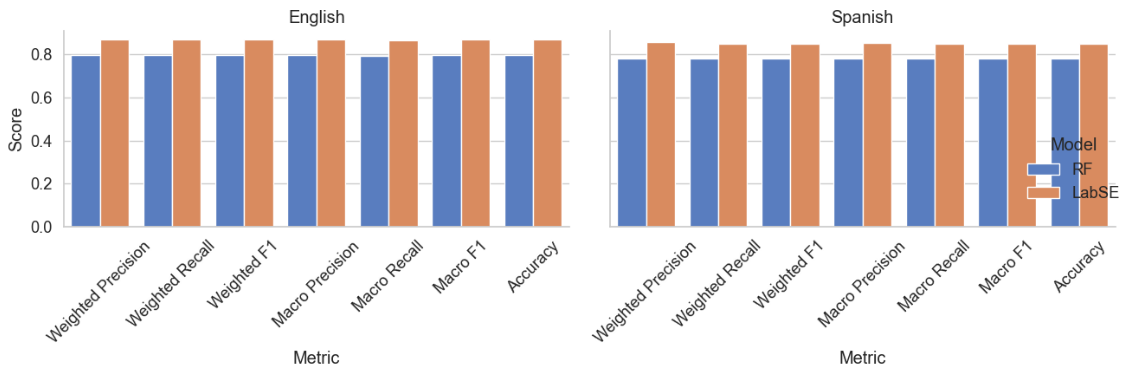

Figure 5: Model comparison by language. LaBSE demonstrates superior generalisation across both English and Spanish, with significant gains over RF.

4.3. Cross-Language and Cross-Model Performance Comparison

The comparative results in Figure 5 and Figure 4 emphasise the effectiveness of LaBSE across languages. Compared to the classical model RF, LaBSE yields substantial improvements, especially in macro F1-score and accuracy. While English benefits slightly more from LaBSE in absolute terms, the improvements in Spanish are equally consistent, reinforcing the model's robustness in multilingual sentiment and intent classification tasks.

4.4. Comparison with Existing Techniques

Recent research in the domain of hope speech detection has explored various models and datasets for multilingual and low-resource classification. Table 4 summarises the key contributions in this area, providing an overview of the reference, dataset, approach, and performance achieved and showing comparism with our approach.

Reference	Dataset	Approach	Best Macro F1 / Accuracy
[22]	English, Spanish	SVM-based classifier	0.489 (English), 0.481 (Spanish)
[23]	English, Spanish	BERT + TF-IDF-based logistic regression	0.82 (binary classification, macro F1)
[25]	English, Urdu, Arabic	BERT-based transfer learning	0.78 (binary F1)
[27]	English, Kannada, Malayalam, Tamil	Ensemble (LSTM, mBERT, XLM-RoBERTa)	0.93 (English), 0.74 (Kannada), 0.82 (Malayalam), 0.60 (Tamil)
[33]	English, Spanish	Transformer vs traditional models (SVM, RF)	0.85 (macro F1, BERT)
[35]/Baseline	English, Spanish	Fine-tuned Transformers (GPT-4, ALBERT)	0.7823 (English, GPT-4 FSL), 0.8425 (Spanish, ALBERT)
Our Work	**English, Spanish**	**LaBSE (Language-agnostic BERT)**	**0.86 (English), 0.83 (Spanish)**

Table 4
Comparison of Existing Techniques with Our Work in Hope Speech Detection

5. Conclusion

In this work, we proposed a hope speech detection technique across languages using the Language-agnostic BERT Sentence Embedding (LaBSE). Our technique tries to minimise the fragmentation that has appeared in previous work by employing a single multilingual model and achieving satisfactory performance on both Spanish and English. In full-fledged testing of the PolyHope IberLEF 2025 test corpus, our technique achieved 86% accuracy on English and 83% on Spanish, outperforming some of the latest models in the literature.

Besides, we have demonstrated the efficiency of LaBSE in handling multilingual tasks without language pipelines, which significantly reduces the burden of model development and maintenance. This approach has been found successful for hope speech detection in structurally diverse languages, and it can potentially be used more extensively across multilingual sentiment analysis tasks.

In the competition, our model ranked second in Spanish and sixth in English, further attesting to the robustness of our method compared to other state-of-the-art methods. This work paves the way for future improvements in cross-lingual classification tasks, particularly in underrepresented languages, where applying pretrained multilingual models like LaBSE would have a significant effect on performance.

Overall, this work contributes to the growing research in multilingual sentiment and emotion analysis, i.e., hope speech, by providing a scalable and efficient method with high potential for real-world usage such as social media monitoring and emotional content analysis.

6. Limitations

Even though our method shows promising results, there are some limitations that need to be addressed in the future.

First, although LaBSE demonstrated strong performance on English and Spanish, its performance on other languages such as Urdu, German and more remains to be evaluated. LaBSE's multilinguality is limited by the quality and quantity of training data for a given language, and we are aware that some less-resourced languages may not benefit as much from this approach.

Second, while the strategy effectively reduces the usage of language-specialized models, it remains rooted in a transformer-based architecture, which is computationally expensive. This limitation may impede the scalability of the model in practical applications, especially on low-end devices such as cell phones or edge computing scenarios.

Moreover, the current model is inclined towards detecting hope speech under a specific context, and further work needs to be conducted to adjust it to detect more intricate emotional states or to address the richness of mixed emotions to a single piece of text. This could be very useful in the context of mental health monitoring or content moderation where multiple emotions could coexist at once.

Finally, our assessment was based mainly on a single dataset (PolyHope IberLEF 2025), and it is possible that this data does not represent the entire diversity of real-world data. Subsequent work must test our method on a larger variety of datasets to validate its generalizability across domains and social contexts.

Overcoming these limitations will be essential to further enhancing the model's performance and broadening its applicability to more use cases.

7. Acknowledgments

The work was done with partial support from the Mexican Government through the grant A1-S-47854 of CONACYT, Mexico, grants 20241816, 20241819, and 20240951 of the Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico. The authors thank the CONACYT for the computing resources brought to them through the Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje of the Laboratorio de Supercómputo of the INAOE, Mexico and acknowledge the support of Microsoft through the Microsoft Latin America PhD Award.

Declaration on Generative AI

We disclose that generative AI tools (e.g. LLMs) were used for drafting and language polishing. The authors retain full responsibility for all content, structure, claims, and conclusions. No AI system was employed to generate data, validate results, or replace human judgment. All factual statements, citations, and analyses were verified by the authors.

References

[1] O. Ojo, A. Gelbukh, H. Calvo, O. Adebanji, Performance study of n-grams in the analysis of sentiments, Journal of the Nigerian Society of Physical Sciences (2021) 477–483.

[2] O. Ojo, A. Gelbukh, H. Calvo, O. Adebanji, G. Sidorov, Sentiment detection in economics texts, in: Advances in Computational Intelligence: 19th Mexican International Conference on Artificial Intelligence, MICAI 2020, Mexico City, Mexico, 12–17 October 2020, Proceedings, Part II, Springer-Verlag, Berlin, Heidelberg, 2020, pp. 271–281. doi:10.1007/978-3-030-60887-3_24.

[3] G. Sidorov, et al., Empirical study of machine learning based approach for opinion mining in tweets, in: MICAI 2012. LNCS (LNAI), volume 7629, Springer, Heidelberg, 2013, pp. 1–14. doi:10.1007/978-3-642-37807-2_1.

[4] S. Aroyehun, A. Gelbukh, Aggression detection in social media: using deep neural networks, data augmentation, and pseudo labeling, in: Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), Association for Computational Linguistics, Santa Fe, New Mexico, USA, 2018, pp. 90–97.

[5] N. Ashraf, R. Mustafa, G. Sidorov, A. Gelbukh, Individual vs. group violent threats classification in online discussions, in: Companion of The 2020 Web Conference, Taipei, Taiwan, 2020, pp. 629–633. doi:10.1145/3366424.3385778.

[6] W. Han, H. Chen, A. Gelbukh, A. Zadeh, L. Morency, S. Poria, Bi-bimodal modality fusion for correlation-controlled multimodal sentiment analysis, in: Proceedings of the 2021 International Conference on Multimodal Interaction, Association for Computing Machinery, New York, NY, USA, 2021, pp. 6–15.

[7] T. T. Hoang, O. E. Ojo, O. O. Adebanji, H. Calvo, A. Gelbukh, The combination of bert and data oversampling for answer type prediction, in: Proceedings of the Central Europe Workshop, volume 3119, 2022.

[8] S. Poria, E. Cambria, A. Gelbukh, Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis, in: EMNLP, 2015.
[9] S. H. Muhammad, N. Ousidhoum, I. Abdulmumin, J. P. Wahle, T. Ruas, M. Beloucif, C. de Kock, N. Surange, D. Teodorescu, I. S. Ahmad, D. I. Adelani, A. F. Aji, F. D. M. A. Ali, I. Alimova, V. Araujo, N. Babakov, N. Baes, A.-M. Bucur, A. Bukula, G. Cao, R. T. Cardenas, R. Chevi, C. I. Chukwuneke, A. Ciobotaru, D. Dementieva, M. S. Gadanya, R. Geislinger, B. Gipp, O. Hourrane, O. Ignat, F. I. Lawan, R. Mabuya, R. Mahendra, V. Marivate, A. Piper, A. Panchenko, C. H. P. Ferreira, V. Protasov, S. Rutunda, M. Shrivastava, A. C. Udrea, L. D. A. Wanzare, S. Wu, F. V. Wunderlich, H. M. Zhafran, T. Zhang, Y. Zhou, S. M. Mohammad, Brighter: Bridging the gap in human-annotated textual emotion recognition datasets for 28 languages, 2025. URL: https://arxiv.org/abs/2502.11926. arXiv:2502.11926.
[10] T. O. Abiola, T. A. Bizuneh, F. Uroosa, N. Hafeez, G. Sidorov, O. Kolesnikova, O. E. Ojo, Cic-nlp at genai detection task 1: Advancing multilingual machine-generated text detection, in: Proceedings of the 31st International Conference on Computational Linguistics (COLING), Association for Computational Linguistics, Abu Dhabi, UAE, 2025.
[11] O. Kolesnikova, A. Gelbukh, A study of lexical function detection with word2vec and supervised machine learning, J. Intell. Fuzzy Syst. 39 (2020).
[12] O. E. Ojo, O. O. Adebanji, H. Calvo, A. Gelbukh, A. Feldman, O. B. Shoham, Doctor or ai? efficient neural network for response classification in health consultations, IEEE Access (2024).
[13] O. O. Adebanji, I. Gelbukh, H. Calvo, O. E. Ojo, Sequential models for sentiment analysis: A comparative study, in: Advances in Computational Intelligence, 21st Mexican International Conference on Artificial Intelligence, MICAI 2022, Proceedings, Part II, Springer, Monterrey, Mexico, 2022.
[14] T. O. Abiola, T. A. Bizuneh, O. J. Abiola, T. O. Oladepo, O. E. Ojo, A. O. O., G. Sidorov, O. Kolesnikova, Cic-nlp at genai detection task 1: Leveraging distilbert for detecting machine-generated text in english, in: Proceedings of the 31st International Conference on Computational Linguistics (COLING), Association for Computational Linguistics, Abu Dhabi, UAE, 2025.
[15] G. Sidorov, F. Balouchzahi, L. Ramos, H. Gómez-Adorno, A. Gelbukh, Mind-hope: Multilingual identification of nuanced dimensions of hope, 2024. Unpublished manuscript.
[16] F. Balouchzahi, S. Butt, M. Amjad, L. J. Gonzalez-Gomez, A. G. M. Meque, H. Gomez-Adorno, B. R. Chakravarthi, G. Sidorov, T. Mandl, R. Priyadharshini, H. Ceballos, S. Rajiakodi, Overview of polyhope-m at ranlp: Bridging hope speech detection across multiple languages, in: Proceedings of the 15th International Conference on Recent Advances in Natural Language Processing (RANLP), 2025.
[17] F. Balouchzahi, G. Sidorov, A. Gelbukh, Polyhope: Two-level hope speech detection from tweets, Expert Systems with Applications 225 (2023) 120078. doi:10.1016/j.eswa.2023.120078.
[18] G. Sidorov, F. Balouchzahi, S. Butt, A. Gelbukh, Regret and hope on transformers: An analysis of transformers on regret and hope speech detection datasets, Applied Sciences 13 (2023) 3983. doi:10.3390/app13063983.
[19] D. García-Baena, F. Balouchzahi, S. Butt, M. García-Cumbreras, A. Tonja, J. A. García-Díaz, S. Jiménez-Zafra, Overview of hope at iberlef 2024: Approaching hope speech detection in social media from two perspectives, for equality, diversity and inclusion and as expectations, Procesamiento del Lenguaje Natural 73 (2024) 407–419.
[20] D. García-Baena, M. García-Cumbreras, S. Jiménez-Zafra, J. A. García-Díaz, V. Rafael, Hope speech detection in spanish: The lgtb case, Language Resources and Evaluation (2023) 1–31. doi:10.1007/s10579-023-09670-y.
[21] S. Butt, F. Balouchzahi, J.-Z. S. M. Amjad, M., H. G. Ceballos, G. a. Sidorov, Overview of PolyHope at IberLEF 2025: Optimism, Expectation or Sarcasm?, in: Procesamiento del Lenguaje Natural, 2025.
[22] M. Yigezu, R. Kannan, H. Shashirekha, Svm-based multilingual hope speech detection for social media texts, in: Proceedings of IberLEF 2023, 2023. URL: https://consensus.app/papers/multilingual-hope-speech-detection-using-machine-yigezu-bade/

0d2fbe6e83c9563e86a9bad09515e3fb.
[23] S. Divakaran, K. Girish, H. Shashirekha, Hope on the horizon: Experiments with learning models for hope speech detection in spanish and english, in: Proceedings of IberLEF 2024, 2024. URL: https://consensus.app/papers/hope-on-the-horizon-experiments-with-learning-models-for-divakaran-girish/2dc9277056415ff2a4ca2ff85597c5d1/.
[24] S. Divakaran, K. Girish, H. L. Shashirekha, Hope on the horizon: Experiments with learning models for hope speech detection in spanish and english, in: Proceedings of IberLEF 2024, CEUR Workshop Proceedings, 2024.
[25] M. Ahmad, U. Sardar, F. Humaira, A. Iqra, M. Muzzamil, A. Hmaza, G. Sidorov, I. Batyrshin, Hope speech detection using social media discourse (posi-vox-2024): A transfer learning approach, Journal of Language and Education (2024). URL: https://consensus.app/papers/hope-speech-detection-using-social-media-discourse-ahmad-sardar/226140cab16752438cd6cf5163c299ae/.
[26] M. Ahmad, I. Ameer, W. Sharif, S. Usman, M. Muzamil, A. Hamza, M. Jalal, I. Batyrshin, G. Sidorov, Multilingual hope speech detection from tweets using transfer learning models, Scientific Reports 15 (2025) 9005.
[27] D. Sharma, V. Gupta, V. K. Singh, B. R. Chakravarthi, Stop the hate, spread the hope: An ensemble model for hope speech detection in english and dravidian languages, ACM Transactions on Asian and Low-Resource Language Information Processing (2025). doi:10.1145/3716383.
[28] T. Nath, V. K. Singh, V. Gupta, Bonghope: An annotated corpus for bengali hope speech detection, International Journal of Information Technology (2025).
[29] M. Arif, M. S. Tash, A. Jamshidi, F. Ullah, I. Ameer, J. Kalita, A. Gelbukh, F. Balouchzahi, Analyzing hope speech from psycholinguistic and emotional perspectives, Scientific Reports 14 (2024) 23548. URL: https://doi.org/10.1038/s41598-024-44146-x. doi:10.1038/s41598-024-44146-x.
[30] F. Balouchzahi, S. Butt, M. Amjad, G. Sidorov, A. Gelbukh, Urduhope: Analysis of hope and hopelessness in urdu texts, Knowledge-Based Systems 308 (2024) 112746. URL: https://consensus.app/papers/urduhope-analysis-of-hope-and-hopelessness-in-urdu-texts-balouchzahi-butt/b73cd3954c745a6c921cb33ea1c19bb4/.
[31] J. Armenta-Segura, G. Sidorov, Ometeotl at hope2024@iberlef: Custom bert models for hope speech detection, in: Proceedings of IberLEF 2024, 2024. URL: https://github.com/JesusASmx/ometeotl-Hope2024.
[32] S. ArunaDevi, B. Bharathi, Machine learning based approach for hope speech detection, in: Proceedings of IberLEF 2024, CEUR Workshop Proceedings, 2024.
[33] L. Eyob, T. Yitbarek, A. Naseeb, G. Sidorov, I. Batyrshin, Enhancing hope speech detection on twitter using machine learning and transformer models, in: Proceedings of IberLEF 2024, CEUR Workshop Proceedings, 2024.
[34] N. Junaida, A. Ajees, Cfilt iit bombay@lt-edi-eacl2021: Hope speech detection for equality, diversity, and inclusion using multilingual representation from transformers, in: Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion (LT-EDI), 2021. URL: https://consensus.app/papers/cfilt-iit-bombayltedieacl2021-hope-speech-detection-for-singh-kumar/3ecd0839d3355845a23eb6624ef15cde/.
[35] S. Butt, F. Balouchzahi, A. I. Amjad, M. Amjad, H. G. Ceballos, S. M. Jiménez-Zafra, Optimism, expectation, or sarcasm? multi-class hope speech detection in spanish and english, https://doi.org/10.13140/RG.2.2.19761.90724, 2025. ResearchGate.
[36] J. Á. González-Barba, L. Chiruzzo, S. M. Jiménez-Zafra, Overview of IberLEF 2025: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41st Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS. org, 2025.