International Journal of Combinatorial Optimization Problems and Informatics, 17(2), 2026, 215-221. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2026.v17i2.1279

www.editada.org

Hope Speech Detection: A Comparative Study across Four Languages

Oluwatobi Joseph Abiola, Tolulope Olalekan Abiola,
Obdulia Pichardo Lagunas, Grigori Sidorov*

[1] Federal University Oye-Ekiti, Nigeria
[2] Instituto Politécnico Nacional, Center for Computing Research, Mexico
[3] Instituto Politécnico Nacional, UPIITA, Mexico
[4] Instituto Politécnico Nacional, Center for Computing Research, Mexico

oluwatobiabiola01@gmail.com, tabiola2025@cic.ipn.mx, opichardola@ipn.mx,
sidorov@cic.ipn.mx
*Corresponding author

Abstract. This paper presents a multilingual approach to hope speech detection using both traditional machine learning (Ridge Classifier with TF-IDF) and transformer-based models (LaBSE). We evaluate performance across four languages English, Spanish, Urdu, and German using the PolyHope dataset. Experimental results demonstrate that LaBSE consistently outperforms the Ridge baseline, particularly in low-resource settings like Urdu. Our findings highlight the effectiveness of multilingual transformers in capturing nuanced expressions of hope across diverse linguistic and cultural contexts.

Article Info
Received Dec 2, 2025
Accepted Jan 4, 2026

1 Introduction

Different social media platforms have developed crucial roles as vital communication channels across the globe, where users employ these platforms to display emotions, exchange life experiences, and create online communities. While scientists have extensively researched how to control toxic materials like hate speech, they are now turning their attention toward the constructive equivalent: hope speech. Textual expressions that build optimistic messages, promote social inclusion, and offer encouragement function as hope speech for people who are suffering or oppressed. This communicative practice serves as a foundation for mental health promotion and social cohesion especially during crises such as pandemics, social unrest, and personal trauma.

The automatic detection of hope speech within digital environments enhances inclusivity by generating positive dialogue that counteracts harmful communication, as shown in recent research Yigezu, (2023) and Divakaran, (2024). However, most current models are trained on monolingual datasets and exhibit limited effectiveness in handling diverse linguistic inputs. This limitation is problematic, as social media today is inherently multilingual and often features code-mixed content.

To address this, we propose a multilingual hope speech detection framework powered by the Language-agnostic BERT Sentence Embedding (LaBSE) model. LaBSE excels in cross-lingual tasks through its ability to generate high-quality sentence embeddings in over 100 languages without requiring parallel corpora. As a transformer-based model, LaBSE creates a shared embedding space for multilingual text encoding, facilitating effective cross-lingual transfer learning without the need for language-specific fine-tuning.

Our project investigates hope speech detection across four diverse languages: English, Spanish, German, and Urdu. These languages represent varying syntactic and morphological characteristics as well as broad global speaker populations. Using LaBSE embeddings within a binary classification framework, we develop a scalable model for inclusive hope speech detection.

Additionally, our research contributes to the field by incorporating underrepresented languages such as Urdu and exploring German-language social media within an extended experimental framework.

2 Literature Review

Text detection and classification have become prominent research areas in Natural Language Processing (NLP), with a wide range of approaches explored over time. Traditional machine learning techniques and deep learning methods (Abdullah et al., 2023b; Achamaleh et al., 2025; Guerrero-Rangel et al., 2024) have laid the groundwork for increasingly accurate and robust classifiers (Abiola et al., 2025; Oladepo et al., 2025; Abdullah et al., 2023a) More recently, hope speech detection has emerged as a specialized subfield, driven by the development of multilingual datasets and the adoption of transformer-based models (Sidorov et al., 2023; Sidorov et al., 2024; Balouchzahi, 2023; Balouchzahi et al., 2025a; Balouchzahi, 2025b; García., 2023; García, 2024).

The field of hope speech detection is relatively new and has gained growing interest, particularly within the domain of social media analysis. Most early studies have concentrated on identifying hopeful text content by classifying optimistic, inclusive, and encouraging statements, often in contrast to the well-established body of research focused on hate speech detection.

Initial approaches to hope speech detection were either monolingual or bilingual. For example, Yigezu et al. (2023) developed a binary hope speech classification system using Support Vector Machines (SVM) for English and Spanish, trained on the IberLEF dataset, achieving moderate F1-scores of 0.489 and 0.481, respectively. Similarly, Divakaran et al. (2024) contributed to the IberLEF HOPE shared task by evaluating a combination of machine learning and transformer-based models, such as BERT, on both binary and multiclass hope speech detection for English and Spanish. Their systems, including GUIDE4Hope and Posi-Vox-2024, achieved macro F1-scores of up to 0.82 in these languages.

Ahmad et al. (2024) introduced the Posi-Vox-2024 dataset and demonstrated the effectiveness of multilingual BERT in classifying hope speech across English, Urdu, and Arabic. Their binary classification model achieved an F1-score of 0.78, underscoring the power of transfer learning in managing underrepresented linguistic populations.

In another study, García-Baena et al. (2023) focused exclusively on the Spanish language, constructing the SpanishHopeEDI dataset featuring LGBT-related content and performing baseline evaluations using traditional machine learning methods for hope speech detection.

Urdu, a low-resource language in NLP, was explored in detail by Balouchzahi et al. (2024), who developed a dual-schema framework encompassing both hopeful and hopeless content. They tested logistic regression and transformer models to examine psychological dimensions of hope in non-Western language contexts. Malik et al. (2023) addressed multilingual transfer learning through fine-tuning RoBERTa on English and Russian datasets. Their findings confirmed that shared multilingual embedding spaces can effectively detect hopeful expressions.

Armenta-Segura and Sidorov (2024) constructed BERT-based models trained on domain-specific hope content in both English and Spanish. They incorporated multilingual sentiment embeddings and achieved promising results in hope speech identification. In a related line of work, Junaida and Ajees (2021) tested multilingual embeddings across English, Tamil, and Malayalam using deep learning models, validating their applicability in low-resource and code-switched environments.

Overall, while these studies show encouraging results using both traditional and transformer-based models, most are limited to one or two languages and rely on language-specific architectures. Very few have attempted to build a unified multilingual classifier for structurally diverse languages. This research addresses that gap by using LaBSE as a shared semantic embedding space for binary hope speech classification across four linguistically distinct languages: English, German, Spanish, and Urdu.

3 Methodology

This section details the approach for building a multilingual hope speech detection system across English, German, Spanish, and Urdu. The research design combines two modeling approaches: one based on TF-IDF with Ridge Classifier, and the other leveraging the transformer-based multilingual LaBSE model.

3.1 Dataset Collection and Description

The datasets used in this study were obtained from a multilingual hope speech detection competition hosted on Codabench. The data includes social media posts labeled with binary tags: Hope and Not Hope, across English, Spanish, Urdu, and German languages. Only binary classification is considered throughout this study.

Each dataset was divided into three partitions: training, development, and testing. Analyses included label distribution and average word count, which influence prediction performance. Table 1 summarizes the key statistics.

Table 1: Dataset Overview

Language	Train Size	Dev Size	Train Hope/Not	Dev Hope/Not	Avg Words (Hope / Not)
English	4541	1650	2296 / 2245	834 / 816	31.3 / 34.6
Spanish	10550	3837	5167 / 5383	1879 / 1958	23.5 / 27.7
Urdu	4613	1678	2183 / 2430	794 / 884	33.4 / 255.2
German	11573	4208	4924 / 6649	1790 / 2418	23.6 / 28.6

3.2 Preprocessing Strategy

A unified preprocessing pipeline was applied to all datasets to ensure consistent feature representation across languages. The primary goal was to normalize text while preserving the emotional tone important for hope speech classification.

3.3 Text Normalization

All text was converted to lowercase. Regular expressions removed URLs, Twitter usernames, and non-alphanumeric symbols. This eliminated platform-specific noise such as hashtags and retweets that could interfere with semantic interpretation.

3.4 Tokenization and Word Count Analysis

Tokenization split text into whitespace-separated tokens. Word count differences between Hope and Not Hope entries were analyzed across languages. In English, Spanish, and German, Not Hope samples were longer. Urdu posed a unique challenge, with some Not Hope entries exceeding hundreds of words. These were retained for analysis.

3.5 Label Encoding

Binary labels were encoded as integers: 1 for Hope, and 0 for Not Hope. This encoding was applied across all datasets.

3.6 Cross-Language Uniformity

The preprocessing routine was language-agnostic and did not rely on language detection. It addressed German compound words and Urdu symbols manually. While LaBSE operated on raw text, TF-IDF required token-clean fidelity.

4 Modeling Architecture

4.1 Ridge Classifier with TF-IDF

TF-IDF vectorization was applied to the cleaned text. The Ridge Classifier from scikit-learn was used with default parameters (alpha=1.0). No cross-validation or dimensionality reduction was used. Evaluation was done on the development sets using accuracy, precision, recall, F1-score, and confusion matrices.

4.2 LaBSE-Based Transformer Classifier

The LaBSE model was fine-tuned from the sentence-transformers/LaBSE checkpoint using Hugging Face's Trainer API. Each language was trained separately with the following configuration:

- Epochs: 5,
- Learning Rate: 2e-5,
- Batch Size (Train/Eval): 32 / 64,
- Dropout (hidden and attention): 0.1.

Checkpoints were saved per epoch, and the model with the lowest validation loss was selected for evaluation.

4.3 Evaluation Setup

Development sets served as validation datasets. Evaluation metrics included accuracy, precision, recall, F1-score, and confusion matrices. These were computed using sklearn.metrics. TF-IDF vectors from the test set were used to generate binary predictions with the Ridge Classifier. Outputs were labeled as Hope or Not Hope and saved as predictions.csv. Text inputs were encoded and predicted using LaBSE. Dummy labels were used during dataset preparation. Predictions were converted using argmax and saved as predictions.csv.

4.4 Postprocessing and Consistency

Text normalization (lowercasing, URL removal, whitespace normalization) was applied to test sets. No additional postprocessing like thresholding or ensemble methods was used. The outputs are reproducible and ready for external applications such as dashboards or moderation tools.

5 Results

Table 2 presents the final evaluation results on the official test set for English, Spanish, Urdu, and German. The models compared are TF-IDF + Ridge Classifier and LaBSE Transformer. Metrics include weighted and macro-averaged precision, recall, and F1-scores, along with overall accuracy.

Table 2: Test Set Results

Language	Model	Accuracy	Macro F1	Weighted F1	Macro Recall
English	Ridge	0.805	0.805	0.805	0.805
English	LaBSE	0.861	0.861	0.861	0.861
Spanish	Ridge	0.784	0.783	0.784	0.783
Spanish	LaBSE	0.870	0.868	0.871	0.871
Urdu	Ridge	0.940	0.940	0.940	0.943
Urdu	LaBSE	0.948	0.948	0.948	0.949
German	Ridge	0.819	0.813	0.818	0.809
German	LaBSE	0.839	0.839	0.839	0.841

5.1 Performance Analysis

Across all four languages, the LaBSE Transformer significantly outperformed the Ridge Classifier on the test set. The most notable gains were observed in Spanish (+8.6% accuracy) and German (+2.0%), where LaBSE's multilingual contextual understanding proved especially beneficial in handling morphological richness and nuanced expressions.
Urdu, a low-resource language, saw consistently high performance from both models. However, LaBSE maintained a slight lead, with a macro F1 of 0.948 compared to Ridge's 0.940. This indicates that deep contextual features helped better generalize over Urdu’s annotation patterns and linguistic cues.

In English, while Ridge performed reasonably well (80.5% accuracy), LaBSE still offered a substantial improvement, achieving 86.1% accuracy and balanced macro and weighted F1-scores, confirming its robustness even in high-resource conditions.

5.2 Model Comparison Summary

The research summary as shown in the figures 1 and 2 is that LaBSE not only increased accuracy but also balanced performance across both Hope and Not Hope classes. Ridge, while competitive in simpler cases, suffered in maintaining class balance and failed to capture deeper linguistic patterns in Spanish and German.

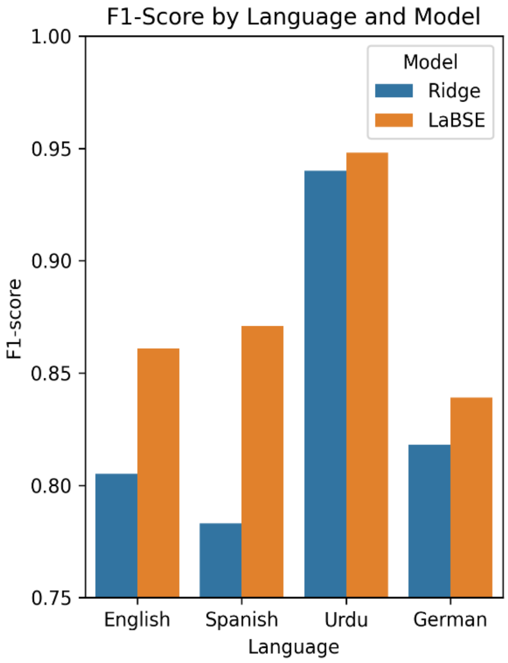

Fig. 1. F1-Score by Language and Model

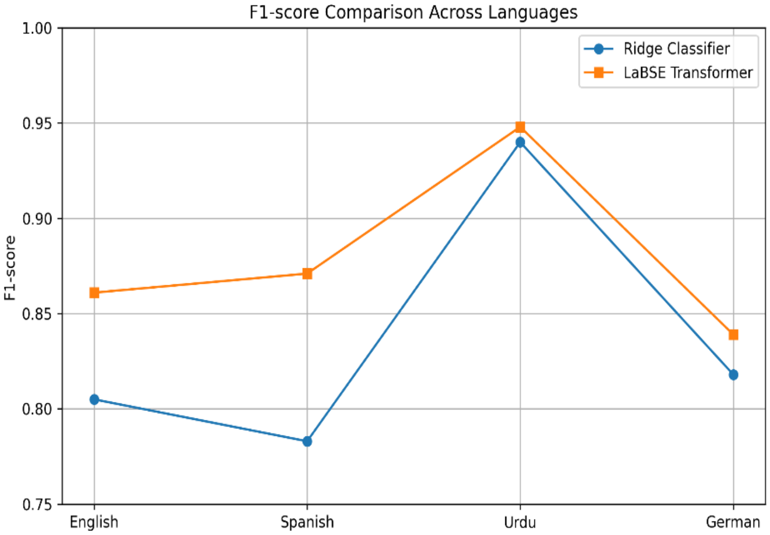

Fig. 2. F1-Score Comparison across Languages

5.3 Comparison with Existing Techniques

As shown in Table 3, earlier approaches to hope speech detection primarily focused on limited language coverage and simpler machine learning models. Yigezu et al. employed Support Vector Machines for English and Spanish using the IberLEF dataset, achieving relatively low macro F1-scores of 0.489 and 0.481.

Table 3: Comparison with Existing Techniques

Reference	Method	Languages	Dataset	F1-score (Macro)
Yigezu et al. (2023)	SVM	English, Spanish	IberLEF	0.489 (En), 0.481 (Es)
Divakaran et al. (2024)	ML + Transformers (BERT)	English, Spanish	IberLEF	Up to 0.82
Ahmad et al. (2024)	mBERT	English, Urdu, Arabic	Posi-Vox-2024	0.78 (avg)
Ours (LaBSE)	LaBSE Transformer	En, Es, Ur, De	PolyHope-M	0.86 (En), 0.87 (Es), 0.95 (Ur), 0.84 (De)

In contrast, Divakaran et al. leveraged BERT-based architectures and reported higher F1-scores (up to 0.82). Ahmad et al. expanded the multilingual scope using mBERT, while our approach achieved significantly higher macro F1-scores across all four languages using LaBSE.

6 Conclusion

This study explored hope speech detection across English, Spanish, Urdu, and German using both a Ridge Classifier and the LaBSE transformer model. While Ridge served as a capable baseline, LaBSE demonstrated superior performance, particularly in Urdu and German, where contextual understanding is essential.
LaBSE's generalizability and robustness highlight its utility for multilingual content moderation.

7 Limitations

Our study has a few limitations. Evaluation relied partly on development data due to the absence of immediate test set labels. The binary classification approach may oversimplify the nuanced nature of hope-related expressions. LaBSE may not fully capture cultural cues in some languages. We also did not assess fairness across demographics or robustness to noisy inputs.

References

Abdullah, A., Hafeez, N., Sardar, K., Oropeza Rodríguez, J. L., Gelbukh, A., & Sidorov, G. (2023). Integration of agile approaches with quantum high-performance computing in healthcare system designs. *Computación y Sistemas, 29*(3). https://doi.org/10.13053/cys-29-3-5810

Abdullah, A., Ullah, F., Hafeez, N., Latif, I., Sidorov, G., Riverón, E. F., & Gelbukh, A. (2023). Cyberbullying detection on social media using machine learning techniques. *Computación y Sistemas, 29*(3). https://doi.org/10.13053/cys-29-3-5481

Abiola, T., Ojo, O. E., Sidorov, G., Kolesnikova, O., & Calvo, H. (2025, July). CIC-IPN at SemEval-2025 Task 11: Transformer-based approach to multi-class emotion. In Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025).

Achamaleh, T., Abiola, T. O., Kawo, L. E., Mebraihtu, M., & Sidorov, G. (2025). CIC-NLP @ DravidianLangTech 2025: Detecting AI-generated product reviews in Dravidian languages. In Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages.

Aggarwal, P., Das, A., & Chakravarthi, B. R. (2023). Multilingual hope-speech detection using transformer-based models. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion (LT-EDI 2023)*. Association for Computational Linguistics. https://aclanthology.org/2023.ltedi-1.38

Ahmad, M., Shahiki-Tash, M., Jamshidi, A., et al. (2024). Analyzing hope speech from psycholinguistic and emotional perspectives. *Scientific Reports, 14*, 23548. https://doi.org/10.1038/s41598-024-74630-y

Ahmad, M., Shahiki-Tash, M., Jamshidi, A., et al. (2024). *Analyzing hope speech from psycholinguistic and emotional perspectives*. *Scientific Reports, 14*, 23548. https://doi.org/10.1038/s41598-024-74630-y

Balouchzahi, F., Sidorov, G., & Gelbukh, A. (2022). *PolyHope: Two-level hope speech detection from tweets* (arXiv:2210.14136). arXiv. https://arxiv.org/abs/2210.14136

Balouchzahi, F., Sidorov, G., & Gelbukh, A. (2023). PolyHope: Two-level hope speech detection from tweets. *Expert Systems with Applications, 225*, 120078. https://doi.org/10.1016/j.eswa.2023.120078

Bruininks, P., & Malle, B. F. (2005). Distinctive features of hope and related emotions. *Cognition & Emotion, 19*(2), 113–142. https://doi.org/10.1080/02699930441000292

Butt, S., Balouchzahi, F., Amjad, A. I., Amjad, M., Ceballos, H. G., & Jiménez-Zafra, S. M. (2025a). *Optimism, expectation, or sarcasm? Multi-class hope-speech detection in Spanish and English* (arXiv:2504.17974). arXiv. https://arxiv.org/abs/2504.17974

Butt, S., Balouchzahi, F., Amjad, M., Jiménez-Zafra, S. M., Ceballos, H. G., & Sidorov, G. (2025b). Overview of PolyHope at IberLEF 2025: Optimism, expectation or sarcasm? *Procesamiento del Lenguaje Natural, 75*, 461–474.

Chakravarthi, B. R. (2020). HopeEDI: A multilingual hope-speech detection dataset for equality, diversity, and inclusion. In *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES 2020)* (pp. 41–53). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.peoples-1.5

Chakravarthi, B. R. (2022). Hope-speech detection in YouTube comments. *Social Network Analysis and Mining, 12*(1), 75. https://doi.org/10.1007/s13278-022-00901-z

Chakravarthi, B. R., & Muralidaran, V. (2021). Findings of the shared task on hope speech detection for equality, diversity, and inclusion. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity* *and Inclusion (LT-EDI 2021)* (pp. 61–72). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.ltedi-1.8

CodaBench. (2025). *PolyHope at IberLEF 2025: Optimism, expectation or sarcasm?* https://www.codabench.org/competitions/5509/

Diener, E. (2009). *The science of well-being: The collected works of Ed Diener*. Springer.

García-Baeza, D., García-Cumbreras, M. Á., Jiménez-Zafra, S. M., García-Díaz, J. A., & Valencia-García, R. (2023). Hope speech detection in Spanish: The LGBT case. *Language Resources and Evaluation, 57*, 1487–1514. https://doi.org/10.1007/s10579-023-09638-3

Guerrero-Rangel, J. R. G., Sidorov, G., Maldonado-Sifuentes, C. E., Vargas-Santiago, M., Ortega-García, M. C., & León-Velasco, D. A. (2024). Natural language processing approach using a neural network ensemble (CNN-HSNN) for skin cancer and multi-disease classification. *Computación y Sistemas, 28*(3). https://doi.org/10.13053/cys-28-3-5015

Jiménez-Zafra, S. M., et al. (2023). Overview of HOPE@IberLEF 2023: Multilingual hope-speech detection. *Procesamiento del Lenguaje Natural, 71*, 289–300. https://doi.org/10.26342/2023-71-29

Khanna, P., Das, A., & Chakravarthi, B. R. (2022). Transformer-based approaches for hope-speech detection. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion (LT-EDI 2022)* (pp. 423–431). Association for Computational Linguistics. https://aclanthology.org/2022.ltedi-1.49

O’Hara, D. J. (2021). Three spheres of hope: Generalised, particularised and transformative. In L. Ortiz & D. O’Hara (Eds.), *Phoenix rising from contemporary global society* (pp. 3–14). Brill.

Oladepo, T., Abiola, O., Abiola, T., Abdullah, A., Muhammad, U., & Abiola, B. (2025, July). *Predicting emotion intensity in text using transformers*. In Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025).

Palakodety, S., KhudaBukhsh, A. R., & Carbonell, J. G. (2019). Hope-speech detection: Helping online communities become more inclusive. In *Proceedings of the 10th ACM Conference on Web Science* (pp. 235–243). Association for Computing Machinery. https://doi.org/10.1145/3292522.3326032

Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N., & Huang, R. (2013). Sarcasm as contrast between a positive sentiment and negative situation. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 704–714). Association for Computational Linguistics.

Sidorov, G., Balouchzahi, F., Ramos, L., Gómez-Adorno, H., & Gelbukh, A. (2025). Multilingual identification of nuanced dimensions of hope speech in social-media texts (MIND-HOPE). *Scientific Reports, 15*(1), 26783. https://doi.org/10.1038/s41598-025-10683-x

Snyder, C. R. (2002). Hope theory: Rainbows in the mind. *Psychological Inquiry, 13*(4), 249–275. https://doi.org/10.1207/S15327965PLI1304_01

Snyder, C. R., Harris, C., Anderson, J. R., et al. (1991). The will and the ways: Development and validation of an individual-differences measure of hope. *Journal of Personality and Social Psychology, 60*(4), 570–585. https://doi.org/10.1037/0022-3514.60.4.570