ISSN 2007-9737

SarcHope: Detection of Sarcasm in Social Media Hope Speech

Girma Yohannis Bade, Grigori Sidorov, Olga Kolesnikova, Jose Luis Oropeza-Rodríguez*

Instituto Politécnico Nacional,
Centro sde Investigación en Computación,
Mexico

girme2005@gmail.com, {sidorov, kolesnikova, joropeza}@cic.ipn.mx

Abstract. Hope speech, which is defined by statements of optimism and encouragement, has a positive impact on social media conversation and is critical in boosting individual well-being, particularly among users who are facing adversity, stress, worry, or illness. As a result, the automatic identification of hopeful content has arisen as an important research topic. However, natural language processing (NLP) systems continue to confront substantial hurdles in accurately detecting hope, which can range from grounded optimism to extreme wishfulness or even scathing sarcasm. To address these issues, we propose SarcHope, a framework for sarcasm detection in hope speech across English and Spanish. Our approach involves re-labeling the IbertLEF-2025 dataset and fine-tuning state-of-the-art transformer-based models. Evaluation on an independent test set shows that DeBERTa-V3 obtains an F1-score of 0.8532 on English data, and S-mmBERT yields a top F1-score of 0.8327 on Spanish data. The results advance the field of natural language processing and provide a valuable baseline for subsequent studies.

Keywords. Hope speech, sarcasm, NLP, transformer-based.

1 Introduction

Social media platforms generate a massive volume of unstructured textual data that is challenging to process yet critically important as it reflects global social interactions and bonding (25). The widespread use of platforms like Twitter and Facebook to express thoughts and emotions presents a ”double-edged sword,” offering social support while also enabling the spread of harmful negativity, such as hate speech. In contrast, the automated detection of supportive content, such as hopeful speech (6), is essential, aiming to foster positive discourses.

Hope is a fundamental human emotion that critically shapes behavior, decision-making, and psychological well-being, offering encouragement during times of stress or isolation (14). Positive expressions of hope speeches have been shown to significantly enhance mental and psychological well-being (40; 37). However, natural language processing (NLP) systems continue to confront substantial hurdles in accurately detecting hopes where intertwined with sarcasm and implicit sentiment, complicating their detection and interpretation (12). Sarcasm refers to words that are the opposite of what people want to say. It becomes a challenge for the application of NLP tasks, including sentiment analysis.

In this study, we propose SarcHope: detecting sarcasm in hope speech by leveraging a dataset from the IberLEF 2025 Shared Task. Our work re-labels the original multiclass of PloyHope IberLEF 2025 data across English and Spanish languages. We condensed the five multiclass labels into three distinct labels as Hope, Not Hope, and Sarcastic Hope. We retrained these data using various state-of-the-art transformer-based models, including BERT, RoBERTa, XLM-R, Deberta-V, and S-mmBERT.The dataset and model details are presented in the subsequent sections.

The outcomes of this study are expected to contribute to more empathetic and context-aware language technologies, with potential applications in mental health monitoring, social media analysis, and sentiment-aware AI systems (5). This work's key contributions include:

1. A theoretical investigation of the concept of hope speech and its treatment with NLP methods.

2. Analyze existing solutions and discuss resulting challenges.

3. Conduct benchmark experiments with five transformer-based models.

4. Analyze errors to establish the study's future paths.

2 Literature Review

2.1 Related Works

The concepts of well-being, thriving, happiness, and life satisfaction reflect the universal emotions shaped by our perceptions of life quality (31). Positive life experiences can enhance personal growth and success, fostering progress across different life stages (39). To further strengthen these benefits, numerous studies have been attempted (13; 14; 38; 11; 34).

Furthermore, (29) introduced a multilingual dataset of hope speeches in English, Tamil, Malayalam, and Kannada aimed at promoting Equality, Diversity, and Inclusion (EDI). This dataset was compiled to advance EDI in language technology and inspire optimism, incorporating contributions from marginalized groups such as the LGBTQIA+ community, individuals with disabilities, and women in STEM fields (6).

They evaluated the HopeEDI dataset using advanced deep learning and machine learning techniques to establish benchmark systems.

Similarly, (26) focused on identifying and amplifying positive and supportive content across platforms. Using transformer-based models, they classified social media comments in English, Malayalam, and Tamil as either hope speech or non-hope speech. Their approach involved transfer learning with leading transformer models for each language. Their results showed ULMFiT achieving an F1-score of 0.9356 for English, while mBERT scored 0.8545 for Malayalam, and distilmBERT reached a weighted F1-score of 0.5926 for Tamil.

In a related study, (21) explored deep learning approaches for word and document representation, employing context-aware string embeddings along with RNN-based text representations. The study evaluated three models across multiple languages using diverse methodologies. Findings indicated that the proposed framework surpassed baseline performance, with Malayalam, Tamil, and English achieving the highest weighted average F-scores of 0.84, 0.93, and 0.58, respectively.

(27) propose using classic machine learning models and the Transformers-Base models on a previously split Hope speech dataset as a training, development, and test set. During testing, a linear-kernel SVM with logistic regression achieved a macro-F1 of 0.78, while an SVM with RBF kernel achieved 0.77 and Naïve Bayes at 0.75. Transformer models performed better, with the top model achieving a weighted precision of 0.82, a weighted recall of 0.80, a weighted F1 of 0.79, a macro F1 of 0.79, and an accuracy of 0.80.

(32) presents an ensemble approach for recognizing hope speech in low-resource languages. Data for four different languages, namely English, Kannada, Malayalam, and Tamil, are collected and tested using several deep learning-based models. An ensemble model is presented to incorporate the benefits of the top-performing models. The proposed Ensemble (LSTM, mBERT, XLM-RoBERTa) model outperforms individual models in all four languages (weighted average F1-scores for English is 0.93, Kannada is 0.74, Malayalam is 0.82, and Tamil is 0.60).

(1) produced a multilingual dataset in English and Urdu, used a translation-based methodology to manage multilingual difficulties, and benchmarked the dataset using a variety of cutting-edge machine learning, deep learning, and transfer learning techniques. Their findings show that a rigorous annotator selection process, combined with specific annotation standards, considerably increased the dataset's quality. Following significant experimentation, their proposed methodology, based on the Bert transformer model, outperformed traditional machine learning models with accuracies of 87% for English and 79% for Urdu. These results demonstrate 8.75% improvement in English and 1.87% in Urdu compared to baseline models (SVM 80% English and 78% Urdu). (28) propose HopeCap, a new multilingual hope speech detection framework. The proposed CapsuleNet takes advantage of the combined

representation of the prediction vector from the child capsule and the final vector produced through dynamic routing. The proposed method calculates three categorization probabilities for the comment's translated, transliterated original script, and transliterated Roman script versions. The terms "original script" relate to the Tamil, Malayalam, and Kannada scripts for their respective languages. The study sheds information on the proposed approach's effectiveness by conducting a systematic examination of HopeCap on three low-resource Dravidian languages. HopeCap surpasses current state-of-the-art approaches by an average of 6.13%, 6.58%, and 4.26% in terms of weighted-F1 for Tamil, Malayalam, and Kannada, respectively. (10) introduced a semi-supervised annotation technique that used Large Language Models (LLMs) and human annotators to annotate the dataset, and they investigated several learning algorithms for hope speech recognition, such as standard machine learning models, neural networks, and transformers.

The hope speech recognition job was divided into two subtasks: a binary classification of Urdu tweets as Hope or Not Hope, and a multiclass classification of Urdu tweets as Generalized, Realistic, and Unrealistic Hopes, as well as Hopelessness and Not Hope (Neutral) categories. Logistic Regression (LR) had the greatest results for binary classification, with an averaged macro F1 score of 0.7593. Transformers outperformed other experiments in multiclass classification, with an averaged macro F1 score of 0.4801.

2.2 Sarcasm in Hope Speech

The presence of sarcasm makes it difficult to detect hopeful speech. Sarcasm in rhetorical expression conveys negative sentiments with a funny or mocking tone. Suppose you say, "This one needs to go well as it happened previously". Ironic discourse disguises semantic pessimism in a seemingly optimistic phrase. Sarcasmic utterances that mimic hopeful statements can cause the model to fail in detecting them. Improper evaluation of sentiment features happens when basic analysis approaches, like positive polarity measurements, are used without extensive contextual analysis (12).

Sarcasm detection is difficult because it requires contextual information, tone signals, and background knowledge. Sarcasm identification in natural language processing has been studied in four different ways: (i) Contextual embeddings (e.g., BERT (2), RoBERTa (22)), (ii) Multitask learning frameworks (e.g., to concurrently learn emotion and sarcasm) (30), (iii) Contrastive learning approaches (20), and (iv) Discourse-aware models (3) that incorporate surrounding sentences or conversation history.

However, sarcasm-aware hope detection remains an unexplored, or no known hope detection dataset includes annotated sarcastic examples, as noted by (11) in their analysis of NPMDU-related textual data. To overcome this issue, we re-labeled hope expressions alongside sarcasm cues to provide a more complete, realistic picture of hope in social conversation.

2.3 Sarcasm and Sentiment Analysis

Sentiment analysis is a technique that analyzes and categorizes people's thoughts, sentiments, and emotions as positive, negative, or neutral.

There are numerous ways for people to convey their sentiments and emotions. Sarcasm is one way to accompany these feelings, particularly when portraying strong emotions (4). It can be described as a positive sentence with a negative intention. Most current research treats sarcasm and sentiment analysis as two distinct tasks, and approaches them primarily as a text categorization problem. However, these approaches cannot appropriately categorize sarcastic statements as negative. To address this issue, the study (33) proposes a multi-task learning-based framework that uses a deep neural network to represent this correlation in order to improve sentiment analysis's overall performance and claims that these two tasks are associated with one another.

Sarcasm is one of the most challenging issues in sentiment analysis. It is commonly used by Indonesian social media users to criticize specific subjects (23).

The authors suggest two new features for detecting sarcasm following sentiment analysis's positive label. The features include negative information and the frequency of interjection words. We used translated SentiWordNet for sentiment categorization.

According to (17), sarcasm is a major challenge for sentiment analysis algorithms. Its complexity stems from the expression of viewpoint through implicit indirect language.

In their research, they propose ArSarcasm, an Arabic sarcasm detection dataset constructed by reannotating existing Arabic emotion analysis datasets. The dataset includes 10,547 texts, of which 16% are sarcastic. In addition to sarcasm, the data was analyzed for sentiment and dialect.

The analysis demonstrates that the very subjective character of these tasks are evidenced by the shift in sentiment labels due to annotators' biases. Experiments reveal that when confronted with sardonic content, cutting-edge sentiment analysers degrade.

3 Methodology

3.1 Datasets

A dataset is crucial for experimental research. This study leverages the dataset from the PolyHope shared task at IberLEF 2025 (18). First, although initially the data features dual label schemes (binary and multiclass), our investigation is dedicated to the multiclass framework. This focus is motivated by the relative abundance of prior work on binary hope speech identification contrasted with the limited exploration of its nuanced categorical variants. We preprocessed the dataset by removing the binary annotation column, resulting in a refined dataframe with two columns: text and multiclass label. This transformation directs the analytical focus toward distinguishing among multiple hope-related categories. Furthermore, the original five-class annotation scheme of multiclass was streamlined to three classes to address the conceptual overlap among hope subcategories. By combining Generalized, Realistic, and Unrealistic Hope into a single Hope class, we reduce label ambiguity and sharpen classification boundaries. Our revised class definitions are:

— **Hope**: Expectations, beliefs, or desires grounded in plausible, meaningful outcomes.

— **Not Hope**: Tweets that do not convey hope, expectation, or desire.

— **Sarcasm**: Texts that might or might not outwardly express hope but are, in fact, sarcastic in nature.

Table 1 shows sample instances of these classes with the corresponding languages.

The data was initially partitioned into two distinct subsets: Training and Test.

Training Data:- This is the subset of data used to train the model, providing both input and corresponding output for learning. The model acquires knowledge by analyzing these real examples (38).

Testing Data:- Once training is complete, the testing data provides an unbiased evaluation of the model's performance. In our study, the model predicts or classifies outcomes based on unlabeled, unseen inputs from this set. We then compare its predictions against manually verified ground-truth outputs to assess accuracy (41).

Table 2 demonstrates an overall view and statistics of the datasets.

3.2 Preprocessing

Text preprocessing serves as a critical foundation in natural language processing, transforming unstructured raw text into a structured format suitable for algorithms to learn from (35; 24). This phase addresses the inherent challenges of real-world textual data, which typically contains noise, inconsistencies, and irrelevant elements that can significantly impair model performance (36; 8). In our implementation, we employ a comprehensive preprocessing pipeline that extends beyond basic text normalization to ensure optimal data quality.

This includes the systematic removal of extraneous characters, numerical values, and special symbols through pattern-based filtering using Python's regex library, which effectively eliminates non-linguistic noise such as URLs and social media artifacts (7). We also standardize

Table 1. Example instances of binary hope speech from the dataset

Class	English	Spanish
Not Hope	Hoping to get on the brown a lot more after I move to the part of the city with a lot of bars.	El deseo una vez saciado se va. El amor no. Cuanto más lo alimentas, más crece. Permanece.
Hope	Pray and hope for nothing but positive energy and vibes to come this way.	Desearía poder estar bien contigo, poder continuar, pero sobre todo, desearía que estemos bien juntos.
Sarcasm	How about this explanation - you're reading WAAAAAY too much into your precious Bible.	Bueno, al menos no puedes ser atacado con una espada desde un coche en movimiento.

Table 2. Data statistics of both languages

Language	Train data	Test data	Total
English	5,233	1,902	7,135
Spanish	11,243	4,088	15,331

whitespace and address other common textual irregularities.

These meticulous cleaning procedures not only enhance dataset coherence but also directly improve model efficacy by reducing vocabulary sparsity, minimizing spurious correlations, and strengthening the signal-to-noise ratio for subsequent feature extraction (7).

This rigorous preprocessing framework ensures that our models operate on high-quality textual representations, which is particularly crucial for both our binary hope-speech classification task and the more complex poly-hope detection task, which must discern subtle nuances, including sarcasm. The attention to preprocessing detail ultimately supports more accurate and reliable model performance across all our classification tasks (9).

3.3 Model

We explore five algorithms (models), including BERT, RoBERTa, XLM-RoBERTa, DeBERTa-V3, and Spanish-mmBERT-small, to retrain our data.

Their descriptions are mentioned below. These models were chosen based on their applicability and previous success with similar NLP issues. All models were fine-tuned with a 2e-5 learning rate and a weight decay of 0.01. We trained in a maximum of three epochs per experiment. The maximum sequence length was set at 128 tokens.

Each investigation were conducted with a batch size of eight for training and sixteen for testing.

1. BERT (Bidirectional Encoder Representation from Transformers):- after the release of the transformer architecture (2), BERT is the first pre-trained LLM model by the Google research team (16). It has two types, BERT-base and BERT-large. For natural language processing (NLP) tasks, including text categorization, sentiment analysis, and question answering, BERT-base is a pre-trained bidirectional language model with 12 layers or blocks, 768 hidden units, and 110M parameters. Both models—BERT-base and large—were pre-trained to optimize Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) for a range of applications. BERT-large is computationally expensive due to its extensive usage of layers and parameters, thus, we used BERT-base in our study.

2. RoBERTa (A Robustly Optimized BERT Pretraining Approach):- For better performance, RoBERTa-base, an optimized variant of BERT-base with 12 layers, 768 hidden units, and 125M parameters, was trained on additional data and did not use Next Sentence Prediction (NSP) (22). Text categorization, sentiment analysis, and question answering are among the NLP tasks where it excels because of its use of dynamic masking and bigger batch sizes.

3. XLM-RoBERTa or XLM-R (Cross-Language Model RoBERTa):- XLM-RoBERTa represents an advanced multilingual variant of RoBERTa, specifically optimized for cross-lingual natural language processing. Unlike standard RoBERTa, this version was pretrained on a substantially larger corpus encompassing over 100 languages through Masked Language Modeling (MLM), while intentionally omitting Next Sentence Prediction (NSP) during training. With its 270 million parameters (15), the model demonstrates superior performance across various multilingual NLP applications, including named entity recognition, sentiment analysis, and machine translation. Notably, empirical evaluations have shown that XLM-RoBERTa-base consistently outperforms multilingual BERT (mBERT) on multiple cross-lingual benchmarks, establishing it as a more robust choice for language-agnostic tasks.

4. Deberta-v3 (Decoding-enhanced BERT with disentangled attention version 3). It's a highly advanced transformer architecture developed by Microsoft that builds upon the original BERT foundation with several key innovations (19). The most notable improvement is its disentangled attention mechanism, which separately represents each word using both content and positional embeddings, allowing the model to better understand relationships between words regardless of their distance in text. DeBERTa-v3 also incorporates an enhanced mask decoder that uses absolute word positions to improve masked token prediction. For pretraining, it employs Replaced Token Detection (RTD) with gradient-disentangled training, making it more sample-efficient than traditional masked language modeling. These architectural advancements have made DeBERTa-v3 a state-of-the-art model particularly excelling in natural language understanding tasks, often outperforming both BERT and RoBERTa variants on benchmarks like GLUE and SuperGLUE, despite having comparable parameter counts.

Table 3. The Pre-trained Transformer Models Used for Both Spanish and English Data

Architecture	Model Name
XLM-R	FacebookAI/xlm-roberta-base
RoBERTa	roberta-base
BERT	bert-base-uncased
DeBERTa-V3	microsoft/deberta-v3-small
S-mmBERT	mrm8488/spanish-mmBERT-small

Table 4. Performance Comparison of Transformer Models on the **English** dataset

		(Weighted Avg)		
Model	**Acc**	**Prec**	**Rec**	**F1**
XLM-R	0.8475	0.8474	0.8475	0.8468
RoBERTa	0.8523	0.8527	0.8523	0.8524
BERT	0.8097	0.8194	0.8097	0.8097
Deberta-v3	**0.8538**	**0.8553**	**0.8538**	**0.8532**
S-mmBERT	0.8475	0.8480	0.8475	0.8477

5. Spanish-mmBERT-small (Spanish multilingual mBERT):-This pruned model performs similarly to the original model for Spanish language tasks, but with a significantly smaller memory footprint [1]. However, because terms not typically used in Spanish were eliminated from the original multilingual model's lexicon, it may underperform in languages other than Spanish.

Table 3 summarizes the architecture of these models, along with their specific names.

4 Result and Discussion

In this section, we evaluate the performance of models using a range of standard classification metrics. Tables 4 and 5 provide average-weighted precision, recall, F1-score, and total accuracy for comparing techniques, tasks, and languages.

From Table 4, it can be observed that all transformer-based models performed well in the English language. Deberta-V3 outperformed others, achieving the weighted average F1-score of 0.8532.

This represents an improvement of +0.044 (4.4%) over BERT, which achieved an F1 score of 0.8097

[1] https://huggingface.co/mrm8488/spanish-mmBERT-small

Table 5. Performance Comparison of Transformer Models on the **Spanish** dataset

		(Weighted Avg)		
Model	**Acc**	**Prec**	**Rec**	**F1**
XLM-R	0.8058	0.8067	0.8058	0.8058
RoBERTa	0.7896	0.7945	0.7896	0.7893
BERT	0.7933	0.8009	0.7933	0.7928
Deberta-v3	0.7999	0.8004	0.7999	0.7998
S-mmBERT	**0.8329**	**0.8347**	**0.8329**	**0.8327**

and was the lowest performer among the models in this category.

This demonstrates that the model's improvement is attributed to its disentangled attention mechanism, which utilizes both content and positional embeddings to represent each word, thereby enhancing the understanding of relationships. Therefore, for the SarcHope task, the Deberta-V3 is the benchmarking model.

Similarly, as shown in Table 5, S-mmBERT attains the top F1-score of 0.8327, surpassing the other models. In this case, RoBERTa's performance declines, resulting in it being the lowest-performing model, with an F1-score of 0.7893.

This shows the largest difference (+0.0434 or 4.34%) when compared to S-mmBERT. This gap may be attributed to S-mmBERT's better adaptation to Spanish data, although it also achieved comparable results on the English dataset. Thus, S-mmBERT is the recommended benchmark for Spanish in the SarcHope task.

The comparison across the languages, for instance, the Deberta-v3 performance in the English data is 0.8532, whereas the Deberta-v3 performance in the English data is 0.7998. The variation is ± 0.0534.

This significant difference might stem from the language-to-model adaptation while uniformly treating the task. On the other hand, S-mmBERT achieved the highest performance, with F1-scores of 0.8327 in Spanish and 0.8477 in English, showing a small difference of ±0.015, indicating the model's tolerance and strength on the language differences.

5 Error Analysis

We use a confusion matrix to analyze errors. This helps us know the specific instances that are correctly and incorrectly classified. In our analysis, we focus on the benchmark models for each language.

Figures 1 and 2, and Tables 6 and 7 show the confusion matrix heatmaps of DeBERTa-V3 and S-mmBERT for the English and Spanish datasets, respectively.

From Figure 1 and Table 6 we can infer the following.

For the Hope class, the model correctly identified 746 Hope instances. However, it misclassified 83 Hope samples as "Not Hope" and 5 as "Sarcastic Hope," yielding a Hope class accuracy of 89.4%.

For the Not Hope Class, 647 samples were correctly predicted. The most errors were made or 149 false positives for Hope (reflecting a tendency toward over-predicting hope), along with 20 false positives for Sarcastic Hope, leading to an accuracy of 79.2%.

The Sarcastic Hope class demonstrated strong performance, predicting 231 correct instances, with very few confusions (7 as Hope and 14 as Not Hope). With an accuracy of 91.7%, it is the most accurately classified category.

Similarly, from Figure 2 and Table 7, we can entail the following analysis for Spanish counterpart:-

The Hope class achieved 1473 correct predictions, with 400 misclassifications as Not Hope and 6 as Sarcastic Hope. This yields a class accuracy of 78.4%. In the Not Hope class, the classifier achieved 1696 true positives, with 259 false positives for Hope and 2 for Sarcastic Hope.

The resulting class accuracy is 86.6%. Finally, in the Sarcastic Hope category, the model correctly identified 236 instances, with merely 3 errors as Hope and 12 as Not Hope. The resulting class accuracy is 94.0%.

ISSN 2007-9737

Table 6. Confusion Matrix for Hope Classification on best performing model(Deberta-V3) in English data

True/Predicted Label	Predicted: Hope	Predicted: Not Hope	Predicted: Sarcastic Hope
Hope	746	83	5
Not Hope	149	647	20
Sarcastic Hope	7	14	231

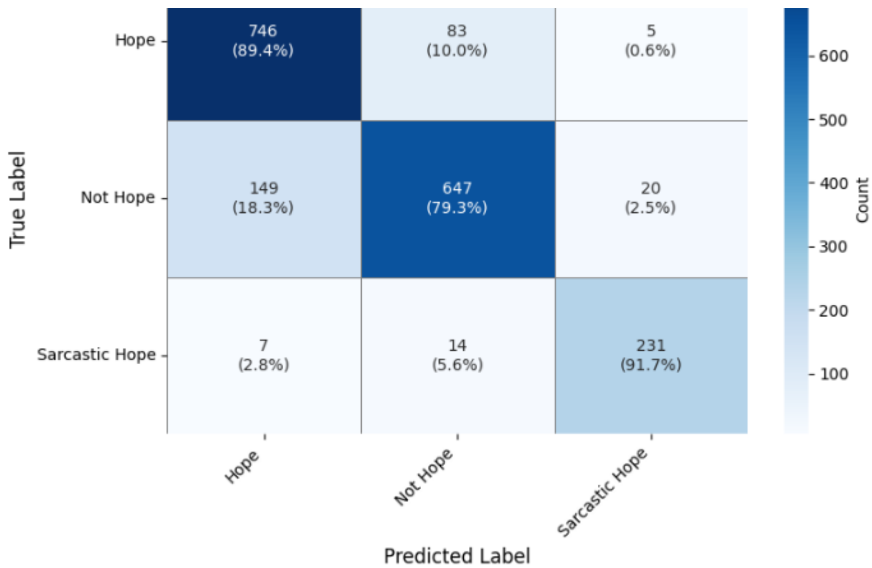

Fig. 1. The figure shows the confusion matrix of the best-performing transformer model (Deberta-V3) in English

6 Conclusion and Future Work

In this study, we figure out a new dimension of hope speech detection known as SarcHope. SarcHope clearly identifies the sarcastic hope from hope and non-hope contents. To achieve this goal, we leverage the dataset from IbertLEF 2025 by re-labeling the classes into the three categories as Hope, Not Hope, and Sarcastic Hope. The dataset comprises the texts in English and Spanish.

To benchmark the task, we explore various transformer-based models, such as BERT, RoBERTa, XLM-R, Deberta-V3, and S-mmBERT. Employing these various models helps us to identify the best benchmarking model and their choice is based on their suitability for the classification task from the previous work. We then retrain these models on 75% of the training dataset, equally treating each model. Following the training phase, evaluation was conducted on an independent test set (25% of the data). Performance was measured using standard classification metrics: accuracy, precision, recall, and F1-score.

To ensure fairness across classes, weighted averaging was applied to precision, recall, and F1-score; accuracy was measured as the total proportion of correctly classified instances. According to the evaluation results, DeBERTa-V3 achieved the best performance on the English dataset with an F1-score of 0.8532, while S-mmBERT performed strongest on Spanish data with an F1-score of 0.8327.

For future work, we recommend the use of a more balanced multilingual dataset to reduce language-specific biases. Additionally, employing large language models (LLMs) for data augmentation, synthetic training examples, or fine-tuning could further improve model generalization and performance across both languages.

Table 7. The Table shows the confusion matrix of the best-performing transformer model (S-mmBERT) in Spanish

True/Predicted Label	Predicted: Hope	Predicted: Not Hope	Predicted: Sarcastic Hope
Hope	1473	400	6
Not Hope	259	1696	3
Sarcastic Hope	3	12	236

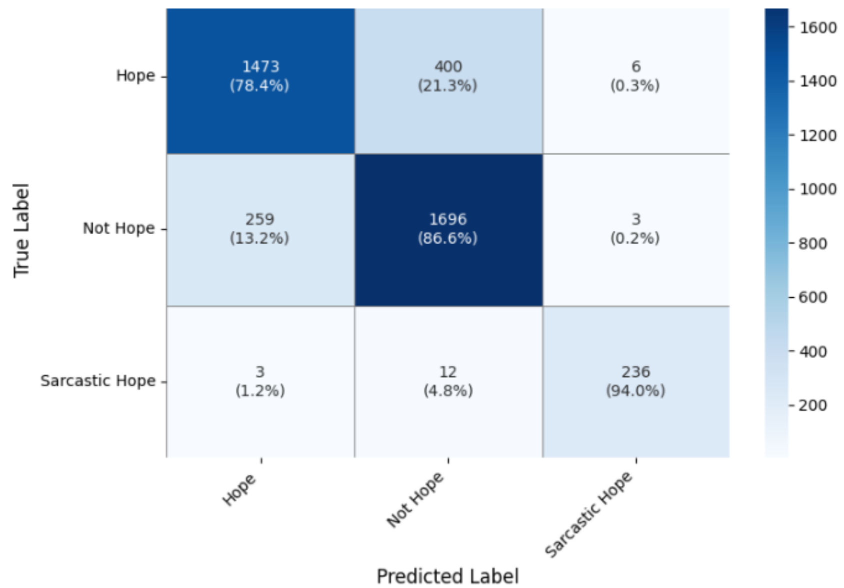

Fig. 2. The figure shows the confusion matrix of the best-performing transformer model (S-mmBERT) in Spanish

Declaration on Generative AI

We used Quill Paraphrase Tool and DeepSeek Generative Model in grammar correction, paraphrasing, and writing refinement. They served as aids for proofreading, rephrasing, and enhancing coherence while preserving the original intent. However, we didn't use them as substitutes for human judgment, particularly in critical, legal, or highly specialized writing.

References

1. **Ahmad, M., Ameer, I., Sharif, W., Usman, S., Muzamil, M., Hamza, A., Jalal, M., Batyrshin, I., Sidorov, G. (2025).** Multilingual hope speech detection from tweets using transfer learning models. Scientific reports, Vol. 15, No. 1, pp. 9005.

2. **Ashish, V. (2017).** Attention is all you need. Advances in neural information processing systems, Vol. 30, pp. I.

3. **Atwell, K., Hassan, S., Alikhani, M. (2022).** Appdia: A discourse-aware transformer-based style transfer model for offensive social media conversations. arXiv preprint arXiv:2209.08207.

4. **Bade, G., Kolesnikova, O., Oropeza, J., Sidorov, G., Yigezu, M. (2025).** Amado at semeval-2025 task 11: Multi-label emotion detection in amharic and english data. Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025), pp. 1406–1410.

5. **Bade, G., Kolesnikova, O., Sidorov, G., Oropeza, J. (2024).** Social media fake news classification using machine learning algorithm. Proceedings of the Fourth Workshop on

Speech, Vision, and Language Technologies for Dravidian Languages, pp. 24–29.

6. **Bade, G. Y., Koleniskova, O., Oropeza, J. L., Sidorov, G., Bergene, K. F. (2024).** Hope speech in social media texts using transformer. Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024), co-located with the 40th Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), CEURWS. org.
7. **Bade, G. Y., Kolesnikova, O., Oropeza, J. L., Sidorov, G. (2024).** Lexicon-based language relatedness analysis. Procedia Computer Science, Vol. 244, pp. 268–277.
8. **Bade, G. Y., Seid, H. (2018).** Development of longest-match based stemmer for texts of wolaita language. International Journal on Data Science and Technology, Vol. 4, No. 3, pp. 79–83.
9. **Bade, G. Y., Zamir, M. T., Kolesnikova, O., Oropeza, J. L., Sidorov, G., Gelbukh, A. (2025).** Girma@ dravidianlangtech 2025: Detecting ai generated product reviews. Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pp. 133–138.
10. **Balouchzahi, F., Butt, S., Amjad, M., Sidorov, G., Gelbukh, A. (2025).** Urduhope: Analysis of hope and hopelessness in urdu texts. Knowledge-Based Systems, Vol. 308, pp. 112746.
11. **Balouchzahi, F., Sidorov, G., Gelbukh, A. (2023).** Polyhope: Two-level hope speech detection from tweets. Expert Systems with Applications, Vol. 225, pp. 120078.
12. **Butt, S., Balouchzahi, F., Amjad, A. I., Amjad, M., Ceballos, H. G., Jimenez-Zafra, S. M. (2025).** Optimism, expectation, or sarcasm? multi-class hope speech detection in spanish and english. arXiv preprint arXiv:2504.17974.
13. **Chakravarthi, B. R. (2020).** HopeEDI: A multilingual hope speech detection dataset for equality, diversity, and inclusion. **Nissim, M., Patti, V., Plank, B., Durmus, E.**, editors, Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media, Association for Computational Linguistics, Barcelona, Spain (Online), pp. 41–53.
14. **Chakravarthi, B. R. (2022).** Hope speech detection in youtube comments. Social Network Analysis and Mining, Vol. 12, No. 1, pp. 75.
15. **Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V. (2019).** Unsupervised cross-lingual representation learning at scale. arXiv preprint arXiv:1911.02116.
16. **Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2019).** Bert: Pre-training of deep bidirectional transformers for language understanding. 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), pp. 4171–4186.
17. **Farha, I. A., Magdy, W. (2020).** From arabic sentiment analysis to sarcasm detection: The arsarcasm dataset. The 4th Workshop on Open-Source Arabic Corpora and Processing Tools, European Language Resources Association (ELRA), pp. 32–39.
18. **González-Barba, J. Á., Chiruzzo, L., Jiménez-Zafra, S. M. (2025).** Overview of IberLEF 2025: Natural Language Processing Challenges for Spanish and other Iberian Languages. 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.
19. **He, P., Gao, J., Chen, W. (2021).** Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. arXiv preprint arXiv:2111.09543.
20. **Hu, H., Wang, X., Zhang, Y., Chen, Q., Guan, Q. (2024).** A comprehensive survey on contrastive learning. Neurocomputing, Vol. 610, pp. 128645.
21. **Junaida, M., Ajees, A. (2021).** Ku_nlp@ lt-edi-eacl2021: a multilingual hope speech detection for equality, diversity, and inclusion using context aware embeddings. Proceedings

of the first workshop on language technology for equality, diversity and inclusion, pp. 79–85.

22. **Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V. (2019).** Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
23. **Lunando, E., Purwarianti, A. (2013).** Indonesian social media sentiment analysis with sarcasm detection. 2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS), IEEE, pp. 195–198.
24. **Mersha, M. A., Yigezu, M. G., Tonja, A. L., Shakil, H., Iskander, S., Kolesnikova, O., Kalita, J. (2025).** Explainable ai: Xai-guided context-aware data augmentation. Expert Systems with Applications, Vol. 289, pp. 128364.
25. **Nath, T., Singh, V. K., Gupta, V. (2025).** Bonghope: An annotated corpus for bengali hope speech detection. International Journal of Information Technology, pp. 1–9.
26. **Puranik, K., Hande, A., Priyadharshini, R., Thavareesan, S., Chakravarthi, B. R. (2021).** Iiitt@ lt-edi-eacl2021-hope speech detection: there is always hope in transformers. arXiv preprint arXiv:2104.09066.
27. **Ramos, L., Calvo, H., Kolesnikova, O. (2025).** Hope speech detection in social media english corpora: Performance of traditional and transformer models. arXiv preprint arXiv:2510.23585.
28. **Rehman, M. Z. U., Raghuvanshi, D., Pachar, H., Raghaw, C. S., Kumar, N. (2025).** Hierarchical attention-enhanced contextual capsulenet for multilingual hope speech detection. Expert Systems with Applications, Vol. 268, pp. 126285.
29. **Roy, P., Bhawal, S., Kumar, A., Chakravarthi, B. R. (2022).** Iiitsurat@ lt-edi-acl2022: Hope speech detection using machine learning. Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, pp. 120–126.
30. **Samant, R. M., Bachute, M. R., Gite, S., Kotecha, K. (2022).** Framework for deep learning-based language models using multi-task learning in natural language understanding: A systematic literature review and future directions. IEEE Access, Vol. 10, pp. 17078–17097.
31. **Shahiki-Tash, M., Armenta-Segura, J., Ahani, Z., Kolesnikova, O., Sidorov, G., Gelbukh, A. (2023).** Lidoma at homomex2023@ iberlef: Hate speech detection towards the mexican spanish-speaking lgbt+ population. the importance of preprocessing before using bert-based models. Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023).
32. **Sharma, D., Gupta, V., Singh, V. K., Chakravarthi, B. R. (2025).** 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.
33. **Tan, Y. Y., Chow, C.-O., Kanesan, J., Chuah, J. H., Lim, Y. (2023).** Sentiment analysis and sarcasm detection using deep multi-task learning. Wireless personal communications, Vol. 129, No. 3, pp. 2213–2237.
34. **Tash, M., Ahani, Z., Zamir, M., Kolesnikova, O., Sidorov, G. (2024).** Lidoma@ lt-edi 2024: Tamil hate speech detection in migration discourse. Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion, pp. 184–189.
35. **Tash, M. S., Ahani, Z., Kolesnikova, O., Sidorov, G. (2024).** Analyzing emotional trends from x platform using senticnet: A comparative analysis with cryptocurrency price. arXiv preprint arXiv:2405.03084.
36. **Tash, M. S., Ahani, Z., Tonja, A., Gemeda, M., Hussain, N., Kolesnikova, O. (2022).** Word level language identification in code-mixed kannada-english texts using traditional machine learning algorithms. Proceedings of the 19th International Conference on Natural Language Processing (ICON): Shared Task on Word Level Language Identification in Code-mixed Kannada-English Texts, pp. 25–28.
37. **Yigezu, M., Kolesnikova, O., Sidorov, G., Gelbukh, A. (2024).** Habesha@

dravidianlangtech 2024: Detecting fake news detection in dravidian languages using deep learning. Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pp. 156–161.

38. **Yigezu, M. G., Bade, G. Y., Kolesnikova, O., Sidorov, G., Gelbukh, A. (2023).** Multilingual hope speech detection using machine learning. CEUR Workshop Proceedings.

39. **Yigezu, M. G., Mehamed, M. A., Kolesnikova, O., Guge, T. K., Gelbukh, A., Sidorov, G. (2023).** Evaluating the effectiveness of hybrid features in fake news detection on social media. 2023 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), IEEE, pp. 171–175.

40. **Yigezu, M. G., Woldeyohannis, M. M., Tonja, A. L. (2021).** Multilingual neural machine translation for low resourced languages: Ometo-english. 2021 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), IEEE, pp. 89–94.

41. **Zamir, M., Tash, M., Ahani, Z., Gelbukh, A., Sidorov, G. (2024).** Tayyab@DravidianLangTech 2024:detecting fake news in Malayalam LSTM approach and challenges. **Chakravarthi, B. R., Priyadharshini, R., Madasamy, A. K., Thavareesan, S., Sherly, E., Nadarajan, R., Ravikiran, M.**, editors, Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, Association for Computational Linguistics, St. Julian's, Malta, pp. 113–118.

Article received on 22/06/2025; accepted on 14/11/2025.
**Corresponding author is Jose Luis Oropeza-Rodríguez.*