NLP-CIC at HOPE2025@IberLEF: Binary and Multi-Class Classification of Hope Speech Detection

Tewodros Achamaleh[1,*,†], Nida Hafeez[1,†], Fatima Uroosa[1,†] and Grigori Sidorov[1,†]

[1]*Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), Mexico City, Mexico*

Abstract

Hope speech supports the development of inclusive digital environments by mitigating hostility and fostering better mental health. We participated in the HOPE task as part of IberLEF 2025 to explore hope speech detection in English and Spanish social media content. The challenge consists of two components: hopeful versus not hopeful classification and multiclass categorization of hope speech forms. Our NLP-CIC team took part in both subtasks of the provided datasets, resulting in an overall placement of 8th place. We applied XLM-roberta-base and BERT-base-multilingual-cased models to perform binary and multiclass classification while adapting it for cross-language hope expression detection. The evaluation showed high scores in the binary classification, where the systems yielded macro F1 scores of 0.8638 for English and 0.8520 for Spanish, and in the multiclass task, the scores reached 0.7139 for English and 0.6901 for Spanish. The results showed that automatic detection of hope is feasible, illustrating its crucial role in fostering positive communication across diverse languages on social platforms. In the binary classification task, our team ranked 8th out of 30 for English and 1st out of 11 for Spanish. For the multiclass classification task, the team achieved 14th out of 28 for English and 4th out of 8 for Spanish

Keywords

Hope Speech Detection, Multilingual NLP, Multiclass Classification, Social Media Text, Machine Learning, Deep Learning, Transformer, XLM-RoBERTa, mBERT, Low-Resource Languages

1. Introduction

Modern social life is deeply interconnected with social networks because these platforms facilitate the sharing of thoughts, opinions and experiences. The expedited growth of social platforms like Twitter and Instagram can be attributed to main characteristics including affordable pricing, limited user identifiability, seamless accessibility and ease of adoption [1]. The NLP network utilizes these social platforms not only as interactive media tools but also as significant data collection repositories [2]. Fundamental NLP tasks such as hope speech detection (HSD) [3], fake news detection (FND) [4], hopeful speech classification (HSC) [5] and sentiment analysis (SA), focus on capturing the complexities of human communication. To handle these tasks, various approaches including machine learning (ML), deep learning (DL) and Transformer based models are employed.

Traditional ML techniques such as logistic regression (LR), support vector machines (SVMs) and naive Bayes (NB) depend on manually engineered features for model building. DL algorithms such as Long Short-Term Memory (LSTM) networks, Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs) have the ability to discover intricate patterns within textual data. The NLP field experienced a significant transformation when Transformer-based models, especially Bidirectional Encoder Representations from Transformers (BERT), embraced attention mechanisms alongside contextual embeddings. The models deliver outstanding achievements when used for different computational tasks [6, 7, 8]. The researchers showed that such models successfully detect both hope and hate speech occurrences on social media platforms [9].

IberLEF 2025, September 2025, Zaragoza, Spain

*Corresponding author.

†These authors contributed equally.

tbizuneh2024@cic.ipn.mx (T. Achamaleh); nhafeez2024@cic.ipn.mx (N. Hafeez); uroosafatima2019@yandex.com (F. Uroosa); sidorov@cic.ipn.mx (G. Sidorov)

0009-0002-9511-0263 (T. Achamaleh); 0009-0008-5490-102X (N. Hafeez); 0000-0002-9575-3157 (F. Uroosa); 0000-0003-3901-3522 (G. Sidorov)

BERT and its various versions lead other models through their exceptional performance and effective cross-lingual capabilities [10]. The massive number of users paired with varied content on social media platforms create unique research opportunities for scientists who want to study human emotions and behavioral patterns and interaction dynamics. Researchers study psychological aspects, emotional states and interpersonal relations alongside belief systems through analysis of user-generated content such as posts, comments, and messages [11]. Social media data enables real-time tracking of social patterns and cultural transitions and public sentiment shifts which make it essential for sociology together with psychology research. Research on hope continues to gain momentum as an emerging focus within this domain, according to [12, 13].

Hope represents the optimistic belief that beneficial results will emerge in the future and functions as a major determinant of mental responses, emotional states, and behavioral patterns [14]. Users of social media platforms regularly post their optimistic visions alongside their ambitions making their expressions suitable for research into this emotion. Future possibilities and possible outcomes that individuals link with hope help shape their mindset even in the face of undefined and challenging situations. Psychological analysis indicates that hope emerges as a positive emotion driven by an individual's expectation of accomplishment of goals, while also revealing the inherent cultural and religious facets of human interaction [15].

The widespread adoption of HSD systems continues to face numerous hurdles, although their usefulness has been established [16]. Advanced preprocessing techniques and filtration methods remain essential for identifying real hope messages among the overwhelming number of social media entries that exist on large scales. The classification results fail to perform accurately because underrepresentation of categories reduces performance but overrepresentation leads to biased outcomes. The detection process for hope outside training boundaries proves difficult since models struggle to extend their capabilities outside their learned contexts. Detecting hope in multilingual settings proves challenging because speakers employ distinct hope language expressions between Spanish and English so specialized detection methods become necessary. Language-aware robust models will be the essential solution to detect hope because they need to effectively process subtle cultural developments.

The workshop organizers launched the HOPE shared task at IberLEF 2025 to develop systems that detect hopefulness across multiple languages. The shared task aims to help researchers develop detection systems that identify hopeful expressions in social media content written in English and Spanish for better interlanguage communication.

2. Related Work

In recent years, hope detection (HD) has gained growing attention for its role in fostering embracing and optimistic communication on social media platforms [17]. For HD, in [18] researchers provide in-depth analysis of different methods through binary text classification into hope and not-hope classes by utilizing SVM and KNN algorithms. Model accuracy improvement depends heavily on hyper parameter optimization because cross-lingual hope detection faces challenges from natural linguistic differences between languages. The research into multilingual classification benefits immensely from these discoveries indicating the necessity of specific algorithm matches for different language contexts. The research field gains significant value from [5] which establishes a new labeled English tweet collection for detecting hope across both binary and multiclass applications. The classification system assigns tweets into two categories: hope and not-hope. Subtypes within hopeful tweets include generalized hope, realistic hope, and unrealistic hope. High levels of agreement between experts were established during careful annotation work that yielded the dataset. Results from experimental testing show standard machine learning algorithms achieve satisfactory binary detection, while transformer models demonstrate better performance during multiclass analysis since they excel at recognizing complex emotional wording. The research urges stakeholders to enhance datasets while studying new platforms and languages to boost the generalization capabilities of models.

The authors in [19] developed a model involving CNN-based architecture during the HOPE 2023 shared task competition at IberLEF. The model analyzes brief texts like English YouTube comments and Spanish tweets through its platform which features a thorough preprocessing mechanism for tokenization with special character management to produce lexical feature outputs. The CNN architectural structure works to discover features which relate to hope. The study identified two main restrictions because mislabeling affected the dataset and conflictive linguistic signals in English text caused class distribution problems. The model proves that CNNs can extract important signals about hope even when working with brief social media messages despite their limitations.

The psycholinguistic characteristics of hope speech receive analysis through established NLP tools, including LIWC, NRC Emotion Lexicon, and VADER sentiment analysis. Different subtypes of hope are explained through linguistic analysis before being classified through LightGBM and CatBoost models. The gradient boosting frameworks deliver effective modeling capabilities for emotional intensity characteristics and hope speech variation through competitive results which surpass traditional classification tools and match deep learning performance [20].

In [21] researchers gives an overview about the IberLEF 2024 workshop, which added value to the NLP community by developing benchmarking tasks and language datasets. Multiple hope speech assessment tasks make use of extensive evaluation metrics for measuring model success. This workshop functions as a key research venue for enhancing language minority studies and encouraging inter-task NLP collaboration among experts. The research highlights how dataset quality combines with model architecture to shape detection of hope trials in terms of linguistic diversity. Research activity regarding the English language predominantly dominates the field even though the Spanish language and others with lower research emphasis need methodological advancements.

This research foundation serves as the basis for our work that builds culturally-sensitive hope detection methods in both English and Spanish languages. Our research utilizes various datasets together with annotation systems to resolve current efficiency constraints while improving hope detection capabilities for both languages.

3. Methodology

We focused on developing a system that could sort short texts within two classification categories ("hope" or "not hope") for binary tasks alongside multiple emotional category identification for multiclass classification. The five categories for multiclass classification included generalized hope, realistic hope and unrealistic hope, as well as not hope with sarcasm. The research aimed to detect different categories of hope by establishing differences between authentic hope and sarcastic expressions. The sentiment analysis task was performed using XLM-R and mBERT transformers, which are pre-trained on extensive multilingual datasets specific to binary classification operations. The model received specialized training to differentiate between texts that expressed hope and those that did not indicate hope. By implementing a specific loss function alongside backpropagation processes, the model learned to adjust its binary classification outputs in alignment with the supplied labeling information. The transformer architecture maintained its design for multiclass classification work but underwent modification to predict between generalized hope, realistic hope, unrealistic hope, not hope, and sarcasm. The prediction of emotional category relied on using categorical cross-entropy as a loss function with softmax activation to determine the most probable outcome.

Both the binary and multiclass tasks utilized k-fold cross-validation to build robustness into the model. The research used five subsets of data divided by K = 5. The training included the use of K-1 subsets, while the final subset functioned as validation in each round. The model underwent multiple epochs of repetition for optimizing its performance outcome. The learning curves showed satisfactory fitting performance in complex datasets which confirmed the capability of our approach to differentiate binary and multiclass emotional categories related to hope.

4. Task Overview

The first task focuses on classifying social media content through a two-part framework between hopeful statements and statements without hope across Spanish and English languages. Online text analysis for hope detection is fundamental to understanding different modes through which hope appears in such content. Identifying hope needs to include both straightforward statements and nuanced indications that might be hidden within complicated language forms and implied emotional content. Subtask **1.a** involves English language binary text classification, while Subtask **1.b** performs a similar Spanish language classification procedure.

The second task separates hope into different categories to identify numerous expressions through a multiclass detection approach. The analysis requires researchers to segregate generalized hope from realistic hope and unrealistic hope together with sarcasm and non-hope within English and Spanish language texts. The multiclass classification approach enables comprehensive exploration of various hope categories to gain full insights into seasonal variations of hope expressed through social media text. Subtask **2.a** and **2.b** apply multiclass classification techniques for English and Spanish language texts to increase analytical details in their respective investigations[22].

4.1. Dataset Analysis

The dataset contains English and Spanish social media texts annotated for both binary and multiclass HSD tasks. The binary classification dataset is almost equally split between "Hope" and "Not Hope" for both languages. The English training set has 2,426 "Hope" and 2,807 "Not Hope" (5,233 total) units, and the development set has 899 "Hope" and 1,003 "Not Hope" (1,902 total) units. Similarly, the training set for Spanish comprises 5,316 "Hope" and 5,927 "Not Hope" samples (11,243 total), and the development set consists of 1,926 and 2,162 samples, respectively (4,088 total). Five categories are used in the multiclass setting: Generalised Hope, Realistic Hope, Unrealistic Hope, Not Hope, and Sarcasm. In this regard, label imbalance becomes more obvious. For English, "Not Hope" dominates the training set (2,245), followed by "Generalized Hope" (1,284), and a high number of "Sarcasm" entries (692). The words "Realistic Hope" and "Unrealistic Hope" are low in frequency (540 and 472, respectively). This trend continues in the development set. The distribution in the Spanish multiclass training data is also similar: The majority class is "Not Hope" (5,383), followed by "Generalized Hope" (2,754), "Realistic Hope", and "Unrealistic Hope" with 1,113 and 1,300 samples, respectively. The lowest class, "Sarcasm", has 693 samples. The pattern for the development set is the same. This disparity, particularly for "Realistic" and "Unrealistic Hope", is a challenge to classification models favouring the majority classes. The more refined "Sarcasm" label brings a new layer of complexity where models must differentiate between the subtle tone and the context. As such, using class-aware training techniques such as weighted loss functions and data augmentation is essential for better model performance, especially in under-represented categories.

Table 1
Dataset Distribution for Binary and Multiclass Hope Speech Detection (English and Spanish)

Language	Set	Generalized Hope	Realistic Hope	Unrealistic Hope	Not Hope	Sarcasm
English	Train	1,284	540	472	2,245	692
English	Dev	467	196	171	816	252
Spanish	Train	2,754	1,113	1,300	5,383	693
Spanish	Dev	1,001	405	473	1,958	251
Language	**Set**	**Hope**	**Not Hope**	**Total**		
English	Train	2,426	2,807	5,233		
English	Dev	899	1,003	1,902		
Spanish	Train	5,316	5,927	11,243		
Spanish	Dev	1,926	2,162	4,088		

The definition of hope in this work aligns with the PolyHope V2 dataset [15], where hope is understood as an emotional state that combines a positive desire with an optimistic expectation of future outcomes. In the binary classification setting, the dataset categorizes tweets as either Hope or Not Hope. In the multiclass setting, it further distinguishes five fine-grained categories: Not Hope, Generalized Hope, Realistic Hope, Unrealistic Hope, and Sarcasm. These categories were applied consistently across both English and Spanish annotations. Table 1 illustrates the dataset distribution across languages.

4.2. System Setup

We used the Hugging Face Transformers library xlm-roberta-base to implement both binary and multiclass classification models for English and Spanish. The preprocessing pipeline consisted of the following steps: (1) text normalization (removing extra spaces, emojis, URLs, and converting to lowercase), (2) label mapping to integer values for classification (e.g., "Generalized Hope" = 0, "Not Hope" = 4), and (3) tokenization using the XLM-RobertaTokenizer, with a maximum sequence length of 256. To address the class imbalance, we computed class weights using inverse frequency from the training data and applied them in a custom loss function during training via the WeightedTrainer class. Data was split using k-fold cross-validation, and each subset was tokenized, padded, and batched using the DataCollatorWithPadding. All code was implemented using PyTorch with mixed precision (FP16) for training efficiency. The ensemble approach combined predictions from XLM-R and mBERT by averaging their output logits before applying the softmax function. This method was chosen over majority voting to preserve the probability distribution across all classes. Each model was fine-tuned separately with identical hyperparameters, and no additional meta-learner was used (i.e., no stacking). The ensemble helped stabilize predictions and improve recall in multiclass classification, especially for rare classes. Early stopping was applied based on macro F1 on the validation set to avoid overfitting.

4.3. Experiments

The training was carried out in five epochs on hardware with a GPU and mixed precision (FP16) for faster training and optimized memory usage. Gradient accumulation was applied to overcome the GPU storage limitations. Early stopping was helped by using the validation macro F1 score to avoid overfitting. For the multiclass setup, we used a batch size of 16 and a cosine scheduler with restarts. For the binary classification setup, the batch size was 8, and the warm-up ratio increased to 0.2. The cosine learning rate schedule was used. The binary setup also applied higher gradient accumulation steps (4) and a lower max gradient norm (0.5). A tokenizer max length of 256 was implemented to normalize the lengths of the sequences. The input sizes within batches were made uniform using the DataCollatorWithPadding to maximize memory and time efficiency. In addition to accuracy and macro F1, we evaluated model performance using ROC (Receiver Operating Characteristic) curves and precision-recall (PR) curves to better assess behavior under class imbalance. These diagnostics helped visualize the trade-off between true positives and false positives, especially for underrepresented categories like Sarcasm and Unrealistic Hope. Curves were generated from validation predictions using Scikit-learn's metrics module. Results showed excellent performance for balanced classification behavior, supported by macro F1 scores. For additional performance improvement, model ensembling was applied, averaging logits from several models. This led to an improved F1 score and more stable predictions. The results, including metrics and visualizations, were saved for reproducibility and further analysis.

5. Results

Binary and multiclass analysis on English and Spanish was carried out with multilingual transformers. XLM-RoBERTa-Base did well in the binary setup, with macro average F1 scores of 0.8638 (English) and 0.8520 (Spanish), implying high reliability in identifying hope speech in the binary setup. The model remained moderately effective for multiclass classification, showing the result of 0.7139 (English) and 0.6901 (Spanish). These results determine the model's good generalization in binary classification, and multiclass models proved more difficult because of label overlaps. Overall, XLM-RoBERTa showed strong multilingual performance across both tasks. Our team in the binary classification task ranked 8^{th} out of 30 for English and 1^{st} out of 11 for Spanish. For the multiclass classification task, the model achieved 14^{th} out of 28 for English and 4^{th} out of 8 for Spanish.

6. Discussion

On both the binary and multiclass classification tasks, XLM-RoBERTa-Base model demonstrated superior performance to mBERT and ensemble models for the English and Spanish tasks. The model produced the best scores on macro F1 in the test set, 0.8638 (binary) and 0.7139 (multiclass) in English and 0.8520 (binary) and 0.6901 (multiclass) in Spanish. These results are a testament to the capacity of XLM-R to capture subtle semantic differences (particularly within fine-grained hope categories). Whether tested on mBERT or not, XLM-R proved to have better recall and class separation, especially in multiclass settings where label overlap is an issue. Class-weighted training improved the detection of underrepresented categories, which the model could generalize well across tasks and languages. The observed lower performance for Spanish, especially in multiclass classification, is likely due to cultural and linguistic nuances in expressing hope, combined with more pronounced class imbalance and fewer semantic cues in the training data. XLM-R benefits English more due to the language's dominance in its pretraining corpus. Nonetheless, XLM-R consistently outperformed mBERT because of deeper architecture, richer multilingual representations, and robustness to class imbalance through class-weighted training strategies. When compared with the PolyHope V2 baseline results [15], our system shows competitive performance. For English binary classification, our XLM-R model achieves a macro F1 of 86.38%, slightly outperforming the baseline RoBERTa score of 86.46%. In the English multiclass task, our macro F1 is 71.39%, closely matching the baseline of 75.87%. In Spanish, our model reaches 85.20% macro F1 (binary) and 69.01% (multiclass), compared to the baseline's 83.97% and 72.81% respectively. These results validate the effectiveness of our multilingual training approach and suggest that XLM-R can achieve near state-of-the-art performance, even without architecture-specific optimization or handcrafted prompt engineering. Tables 2 and 3 compare the models' F1 Scores across all languages for the binary and multiclass settings.

Table 2
Model Comparison for Multiclass Classification

Model	Lang	Task	Accuracy	F1	Precision	Recall
XLM-R Base	en	Multiclass	0.7219	**0.7052**	0.6877	0.7541
mBERT	en	Multiclass	0.6919	0.6716	0.6528	0.7190
Ensemble	en	Multiclass	0.7129	0.6954	0.6768	0.7492
XLM-R Base	es	Multiclass	0.6837	0.6717	0.6442	0.7456
mBERT	es	Multiclass	0.6720	0.6686	0.6499	0.7210
Ensemble	es	Multiclass	0.6898	**0.6824**	0.6567	0.7569

Table 3
Model Comparison for Binary Classification

Model	Lang	Task	Accuracy	F1	Precision	Recall
XLM-R Base	en	Binary	0.8502	**0.8462**	0.8218	0.8721
mBERT	en	Binary	0.8465	0.8387	0.8332	0.8443
Ensemble	en	Binary	0.8454	0.8441	0.8065	0.8854
XLM-R Base	es	Binary	0.8400	**0.8446**	0.7787	0.9226
mBERT	es	Binary	0.8200	0.8207	0.7732	0.8744
Ensemble	es	Binary	0.8383	0.8413	0.7825	0.9097

7. Error Analysis

Substantial misclassification, particularly between closely related semantic categories, was found in both the English and Spanish datasets, indicating the model's difficulty in accurately classifying nuanced expressions of Hope.

7.1. English Binary Classification

In the English binary classification task, the model confused 170 instances of Not Hope as Hope and 115 Hope instances were misclassified as Not Hope. These errors suggest challenges in drawing a clear boundary between neutral or ambiguous language and genuine hopeful expressions.

7.2. English Multiclass Classification

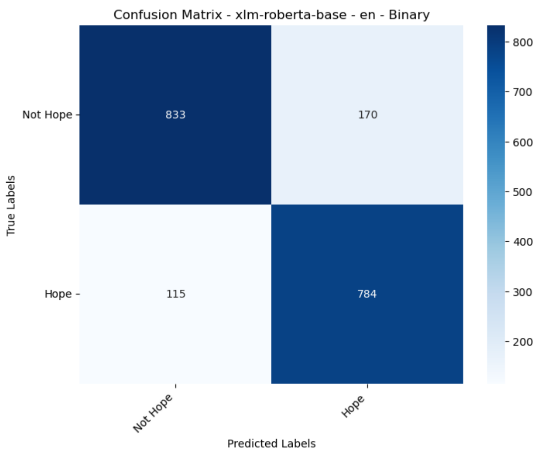
The English multiclass task revealed further confusion: 99 instances of Not Hope were misclassified as Generalized Hope, and 88 as Unrealistic Hope. Additionally, Generalized Hope was often mistaken for Realistic Hope. These confusions highlight the challenge of distinguishing between vague optimism (Generalized Hope), feasible positive expectations (Realistic Hope), and overly idealistic or improbable expressions (Unrealistic Hope), particularly when contextual cues are limited.

7.3. Spanish Binary Classification

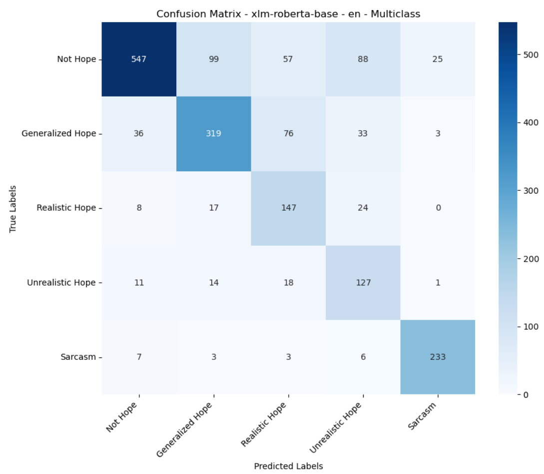
In Spanish, binary classification errors were even more pronounced, with 505 Not Hope texts misclassified as Hope, and 149 in the opposite direction. This suggests a language-specific issue in capturing subtle sentiment cues that indicate hopelessness or lack of positivity.

7.4. Spanish Multiclass Classification

The Spanish multiclass task showed significant overlap between Not Hope and Generalized Hope (247 misclassifications) and Not Hope and Unrealistic Hope (280). There were also 130 confusions between Generalized and Realistic Hope. These patterns reflect the difficulty of modeling culturally embedded expressions and tone, particularly when sarcasm, exaggeration, or indirect emotional references are present. To gain deeper insight into these errors, we performed a qualitative review of misclassified instances. These findings underline the limitations of the current system in handling subtle tone, semantic ambiguity, and context-dependent meaning.

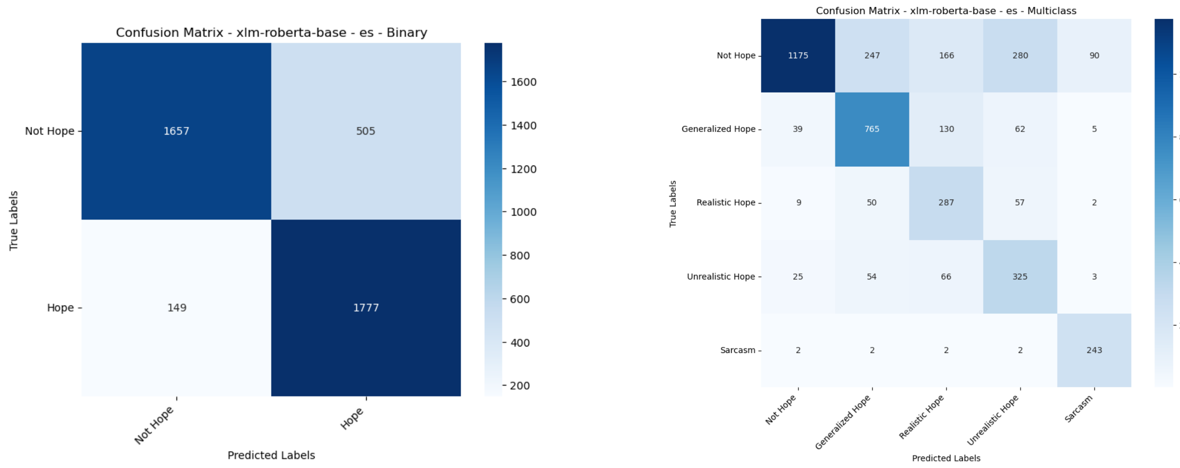


(a) English Binary



(b) English Multiclass

Figure 1: Confusion matrices for English in binary and multiclass classification tasks.



(a) Spanish Binary

(b) Spanish Multiclass

Figure 2: Confusion matrices for Spanish in binary and multiclass classification tasks.

The model often struggles when expressions of hope are implied rather than explicit, or when multiple emotional cues coexist in the same text. Additionally, overlapping definitions between categories (e.g., Generalized vs. Realistic Hope, or Unrealistic Hope vs. Not Hope) increase the likelihood of confusion. Annotation inconsistencies or ambiguous labels in the dataset may further contribute to these errors. Addressing these challenges would benefit from improved semantic representations, more consistent and context-aware annotation guidelines, and the integration of specialized modules such as emotion or sarcasm detection to enhance classification performance in future work. Figure 1a,1b,2a and 2b provides confusion matrices that visualize specific patterns related to misclassification.

8. Conclusion

This study assessed multilingual hope speech detection with XLM-RoBERTa-Base, mBERT, and an ensemble of the two models for English and Spanish for both binary and multiclass settings. Experiments revealed that XLM-RoBERTa-Base stabilised to record optimal performance, especially in distinguishing between fine-grained groups of hope speech. It showed good generalization and robustness across languages and outperformed baseline models in precision and recall. Though F1 scores from binary classification were higher overall, multiclass classification was more complex and emphasized the need for more sophisticated contextual modeling. In the future, more work may be done on covering low-resource languages and adding the cultural idiosyncrasies to improve speech recognition and enhance multilingual speech recognition.

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

The authors acknowledge limited assistance from generative AI for text editing purposes. All conceptualization, experimentation, and interpretation are original human work. Responsibility for the accuracy and integrity of the manuscript rests solely with the authors.

References

[1] D. García-Baena, F. Balouchzahi, S. Butt, M. Á. García-Cumbreras, A. L. Tonja, J. A. García-Díaz, S. Bozkurt, B. R. Chakravarthi, H. G. Ceballos, R. Valencia-García, et al., 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.

[2] S. Butt, N. Ashraf, M. H. F. Siddiqui, G. Sidorov, A. Gelbukh, Transformer-based extractive social media question answering on tweetqa, Computación y Sistemas 25 (2021) 23–32.

[3] D. García-Baena, M. Á. García-Cumbreras, S. M. Jiménez-Zafra, J. A. García-Díaz, R. Valencia-García, Hope speech detection in spanish: The lgbt case, Language Resources and Evaluation 57 (2023) 1487–1514.

[4] T. Achamaleh, N. Hafeez, M. Mebraihtu, F. Uroosa, G. Sidorov, Cic-nlp@dravidianlangtech 2025: Fake news detection in dravidian languages, in: Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, 2025, pp. 647–654.

[5] F. Balouchzahi, G. Sidorov, A. Gelbukh, Polyhope: Two-level hope speech detection from tweets, Expert Systems with Applications 225 (2023) 120078.

[6] T. Achamaleh, T. O. Abiola, L. E. Kawo, M. Mebraihtu, G. Sidorov, 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, 2025, pp. 502–507.

[7] T. O. Abiola, T. A. Bizuneh, O. J. Abiola, T. O. Oladepo, O. E. Ojo, G. Sidorov, O. Kolesnikova, Cic-nlp at genai detection task 1: Leveraging distilbert for detecting machine-generated text in english, in: Proceedings of the 1st Workshop on GenAI Content Detection (GenAIDetect), 2025, pp. 271–277.

[8] 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 1st Workshop on GenAI Content Detection (GenAIDetect), 2025, pp. 262–270.

[9] 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.

[10] J. Armenta-Segura, G. Sidorov, Ometeotl at hope2024@ iberlef: Custom bert models for hope speech detection, in: 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), CEUR-WS. org, 2024.

[11] M. S. Tash, Z. Ahani, O. Kolesnikova, G. Sidorov, Analyzing emotional trends from x platform using senticnet: A comparative analysis with cryptocurrency price, arXiv preprint arXiv:2405.03084 (2024).

[12] S. Butt, F. Balouchzahi, M. Amjad, S. M. Jiménez-Zafra, H. G. Ceballos, G. Sidorov, Overview of polyhope at iberlef 2025: Optimism, expectation or sarcasm?, Procesamiento del Lenguaje Natural (2025).

[13] F. Balouchzahi, S. Butt, M. Amjad, G. Sidorov, A. Gelbukh, Urduhope: Analysis of hope and hopelessness in urdu texts, Knowledge-Based Systems 308 (2025) 112746.

[14] F. Ullah, M. T. Zamir, M. Ahmad, G. Sidorov, A. Gelbukh, Hope: A multilingual approach to identifying positive communication in social media, in: 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), CEUR-WS. org, 2024.

[15] S. Butt, F. Balouchzahi, A. I. Amjad, M. Amjad, H. G. Ceballos, S. M. Jimenez-Zafra, Optimism, expectation, or sarcasm? multi-class hope speech detection in spanish and english, arXiv preprint arXiv:2504.17974 (2025).
[16] B. R. Chakravarthi, Hopeedi: A multilingual hope speech detection dataset for equality, diversity, and inclusion, in: Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media, 2020, pp. 41–53.
[17] M. Krasitskii, O. Kolesnikova, L. C. Hernandez, G. Sidorov, A. Gelbukh, Hope2023@ iberlef: A cross-linguistic exploration of hope speech detection in social media, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEURWS. org, 2023.
[18] Z. Ahani, G. Sidorov, O. Kolesnikova, A. F. Gelbukh, Zavira at hope2023@ iberlef: Hope speech detection from text using tf-idf features and machine learning algorithms., in: IberLEF@ SEPLN, 2023.
[19] M. Tash, J. Armenta-Segura, Z. Ahani, O. Kolesnikova, G. Sidorov, A. Gelbukh, Lidoma@dravidianlangtech: Convolutional neural networks for studying correlation between lexical features and sentiment polarity in tamil and tulu languages, in: Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages, 2023, pp. 180–185.
[20] M. Arif, M. S. Tash, A. Jamshidi, I. Ameer, F. Ullah, J. Kalita, A. Gelbukh, F. Balouchzahi, Exploring multidimensional aspects of hope speech computationally: A psycholinguistic and emotional perspective (2024).
[21] L. Chiruzzo, S. M. Jiménez-Zafra, F. Rangel, Overview of iberlef 2024: Natural language processing challenges for spanish and other iberian languages, in: 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), CEUR-WS.org, 2024.
[22] 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.