https://doi.org/10.17323/jle.2024.22443

Hope Speech Detection Using Social Media Discourse (Posi-Vox-2024): A Transfer Learning Approach

Muhammad Ahmad [1], Sardar Usman [2], Humaira Farid [3], Iqra Ameer [4], Muhammad Muzammil [5], Ameer Hamza [5], Grigori Sidorov [1], Ildar Batyrshin [1]

[1] Instituto Politecnico Nacional (CIC-IPN), Mexico City, Mexico
[2] Institute of Arts and Culture, Lahore, Pakistan
[3] Independent Researcher, California, USA
[4] Pennsylvania State University at Abington, PA, USA
[5] Islamia University of Bahawalpur, Pakistan

ABSTRACT

Background: The notion of hope is characterized as an optimistic expectation or anticipation of favorable outcomes. In the age of extensive social media usage, research has primarily focused on monolingual techniques, and the Urdu and Arabic languages have not been addressed.

Purpose: This study addresses joint multilingual hope speech detection in the Urdu, English, and Arabic languages using a transfer learning paradigm. We developed a new multilingual dataset named Posi-Vox-2024 and employed a joint multilingual technique to design a universal classifier for multilingual dataset. We explored the fine-tuned BERT model, which demonstrated a remarkable performance in capturing semantic and contextual information.

Method: The framework includes (1) preprocessing, (2) data representation using BERT, (3) fine-tuning, and (4) classification of hope speech into binary ('hope' and 'not hope') and multi-class (realistic, unrealistic, and generalized hope) categories.

Results: Our proposed model (BERT) demonstrated benchmark performance to our dataset, achieving 0.78 accuracy in binary classification and 0.66 in multi-class classification, with a 0.04 and 0.08 performance improvement over the baselines (Logistic Regression, in binary class 0.75 and multi class 0.61), respectively.

Conclusion: Our findings will be applied to improve automated systems for detecting and promoting supportive content in English, Arabic and Urdu on social media platforms, fostering positive online discourse. This work sets new benchmarks for multilingual hope speech detection, advancing existing knowledge and enabling future research in underrepresented languages.

KEYWORDS

hope speech, BERT, machine learning, twitter analysis, social media, transfer learning, NLP

Citation: Ahmad M., Sardar U., Humaira F., Iqra A., Muzzamil M., Hmaza A., Sidorov G., & Batyrshin I. (2024). Hope Speech Detection Using Social Media Discourse (Posi-Vox-2024): A Transfer Learning Approach. *Journal of Language and Education, 10*(4), 31-43. https://doi.org/10.17323/jle.2024.22443

Correspondence:
Ildar Batyrshin,
batyr1@cic.ipn.mx

Received: September 7, 2024
Accepted: December 16, 2024
Published: December 30, 2024

INTRODUCTION

Hope is defined as a positive emotional state that includes expectations or anticipation of beneficial outcomes in the future. Many online social media platforms have provided a space for millions of users to voice their thoughts and share their views. This opportunity not only generated negative content but also fostered the exchange of positive ideas (Alawadh et al., 2023) and promoted positivity. Recently, hope speech detection on social media has gained significant attention, with few studies addressing this issue across both high- and low-resource languages (Arif et al., 2024; Balouchzahi et al., 2023; Chakravarthi, 2022). Hope speech detection is relatively new approach that focuses on identifying and amplifying positive online content to promote social harmony and encourage a more positive atmosphere within communities. Among the limited studies on hope speech detection, research has primarily focused on monolingual contexts, developing individual classification models tailored to each language, such as English (Balouchzahi et al., 2023), Spanish (Kumar et al., 2022), English, Tamil

and Malayalam (RamakrishnaIyer et al., 2023), and Bengali (Nath et al. 2023), while Arabic and Urdu languages have not been addressed in either monolingual or multilingual contexts.

For many, social media has become a vital platform for seeking support (Gowen et al., 2012; Yates et al., 2017; Wang & Jurgens, 2018). Social integration is essential for their overall well-being, particularly for those vulnerable to exclusion. By identifying and amplifying encouraging messages on social media, hope speech detection can contribute to a more equitable and inclusive digital landscape. Additionally, the methodology developed in this study has broad applications in psycholinguistics and natural language processing, where it can be used to identify positive sentiment, resilience, and constructive discourse across various contexts.

Social media platforms host numerous hateful or malicious posts (Louati, Ali, et al., 2024; Irfan, Asim, et al., 2024; Anjum, and Rahul Katarya., 2024), largely because of the lack of regulatory authority. Analyzing content on Twitter and other platforms has proven effective in curbing the spread of negativity through techniques like hate speech detection (Schmidt & Wiegand, 2017, Subramanian, Malliga, et al., 2023; Nagar, Barbhuiya, & Dey, 2023), offensive language identification (Anand et al., 2023; Kogilavani et al., 2023; Mnassri et al., 2024), and abusive language detection (Zampieri et al., 2019; Austin et al., 2020; Yenala et al., 2018). Nonetheless, as highlighted by recent research, existing technologies for detecting abusive language (Lee et al., 2018) often fail to account for the potential biases inherent in the datasets upon which they are trained. The presence of systematic racial biases within these datasets can render abusive language detection systems inherently biased, leading to discriminatory outcomes that disproportionately affect minority or marginalized groups. Such biases in language detection technology have the potential to perpetuate discrimination (Davidson et al., 2019). Therefore, we should prioritize promoting positive interactions rather than merely addressing individual negative posts. In this context, hope speech detection offers a novel approach, not only by counteracting negativity but also by contributing to a more positive and inclusive online environment across a wide range of linguistic and cultural contexts. To achieve this objective, we have created a comprehensive joint multilingual hope speech corpus for Urdu, Arabic, and English, using binary and multi classification. The process begins by collecting data related to hope speech tweets in English, Urdu, and Arabic from Twitter. After the collection of the dataset we pre-processed each sample to make it more robust for machine learning models. After pre-processing, the data goes through a joint multilingual process, where the English, Urdu, and Arabic datasets are combined. In the annotation phase, the data is labeled according to specific guidelines. The next step involves fine-tuning our proposed models and applying them to the dataset for classification tasks. Finally, the different machine learning and deep learning and transformer based models are evaluated for accuracy, F1-score, recall, and precision, and the results are analyzed for binary and multi-class classification tasks. This methodology provides a comprehensive approach to hope speech detection across different languages.

This study makes the following contributions:

1. To the best of our knowledge, joint multilingual hope speech detection for English, Urdu, and Arabic has not been developed earlier, and we have explored a comprehensive joint multilingual corpus with extensive guidelines for annotating the dataset;
2. We explored hope speech detection as a two-level text classification task for the first time in joint multilingual dataset (English, Arabic, and Urdu) languages and propose a multiclass classification approach for Urdu and Arabic languages;
3. A comprehensive series of experiments demonstrated that the proposed methodology achieved the best performance compared to the baseline;
4. The proposed framework demonstrated a 0.78 accuracy rate in binary class and a 0.66 accuracy rate in multi-class to our dataset. This represents improvements of 0.04 in accuracy in binary and 0.08 accuracy rate in multi-class compared to the baseline performance metrics.

LITERATURE REVIEW

Existing Datasets for Hope Speech Detection

The process of corpus generation for hope speech detection has become a major focus in this field, although these corpora are typically limited in terms of language coverage and sample size. For example, Balouchzahi et al. (2023) recently introduced a dataset for detecting hope speech in English and applied machine learning, deep learning, and transformer-based methods to benchmark the dataset. However, this dataset was limited to a single language, and the study did not address multilingual classification. Similarly, Chakravarthi (2022) introduced a CNN model for hope speech detection in English and Dravidian languages but did not address multi classification. These studies highlight the need for more diverse datasets, including multiple languages, to improve generalization. Furthermore, Chakravarthi (2022) created a joint multilingual dataset for English, Tamil and Malayalam language using YouTube comment to recognize and encourage positivity in the comments but the author did not use multi-classification task in hope speech detection.

Multilingual Hope Speech Detection

Several studies have explored multilingual hope speech detection, employed advanced machine learning models to handle linguistic and cultural differences across languag-

es. Ghanghor et al. (2021) applied pre-trained transformer models such as m-BERT-cased and XLM-RoBERTa for detecting hope speech in English, Tamil, and Malayalam. Their Results shows that m-BERT-cased perform better than all other models, achieving a highest F1-score of 0.93 for English, 0.83 for Malayalam, and 0.60 for Tamil. While this work contributes to multilingual detection but it does not explore multi classification task across diverse languages. Moreover, Chinnappa (2021) worked on detecting hope speech in Tamil, English, and Malayalam, highlighting the challenges posed by code-mixed data, which further complicates the classification task. Building on this, Malik et al. (2023) extended the scope by exploring a joint multilingual and translation-based approach, focusing on English and Russian languages, highlighting the potential of translation techniques in multilingual hope speech detection. They fine-tuned a pre-trained Russian-RoBERTa model and achieved impressive results, with an accuracy of 94% and an F1-score of 80.24%. This approach demonstrated the potential of leveraging translation for better model performance, but it did not address multiclass classification tasks, which remain an important area for further exploration.

Contribution of the Current Research

Our research presents a novel approach by focusing specifically on joint multilingual hope speech detection across English, Urdu, and Arabic languages using two level text classification. Unlike prior researches that have focused on individual languages, our methodology offers a comprehensive joint-multilingual dataset that comprises both binary and multiclass classification tasks. This study contributes valuable insights into the detection of hope speech across three languages, offering new avenues for improving sentiment analysis and social media monitoring tools. Table 1 provides a summary of prior studies related to hope speech detection, highlighting the differences between these studies and the proposed study.

METHOD

Corpus Compilation Process

Dataset Collection and Integration

Our dataset consists of approximately 80,000 tweets from various disciplines in English, Urdu, and Arabic that were sourced from Twitter[1], as Twitter is the largest social media platform and microblogging service that enables users to post and interact with messages known as «tweets.» The collection process involved extracting tweets using Twitter's API (Tweepy). In this study, we amassed a corpus of 80,000 recent and keywords-based tweets sourced from Twitter, employing a systematic approach centered on hope-related keywords, like in Urdu انشاء الله (In Sha Allah), خیر انشاء الله (Khair In Sha Allah), خواہش ہے (wish), کل (tomorrow), مستقبل (future), کامیابی (success), انتظار (waiting), امید سے بھرپور (hopeful), etc. while in English we used (aspiration, believe, coming soon, dreaming, expectation, feeling positive, I wish, looking forward to and joyful), etc. and while for Arabic we used التفاؤل (optimism), تشجيع (encouragement), ذهاب, موافقة (approval), يتمنى (wish), etc., with different variations. These keywords were used to capture a diverse spectrum of hopeful expressions and sentiments articulated across the

Table 1

Prior Studies Related to Hope Speech Detection vs. Proposed Study

References	Language	Joint Multilingual	Supervised Methods	Multi Classification
Balouchzahi et al. (2023)	English	No	LR, SVM, CNN, LSTM, BiLSTM, Transformer	Yes
Malik et al. (2024)	English, Russian	Yes	SVM, RF, CNN, RoBET base with classifier	No
Kumar et al. (2022)	English, Spanish, Tamil, Malayalam	No	SVM, LR, RF	No
Roy et al. (2022)	English	No		No
Chakravarthi et al. (2021)	English, Tamil, Malayalam, Kannada	No	SVM, DT, LR, KNN, RoBERTa Classifier	No
Ghanghor et al. (2021)	English	No		No
Proposed	English, Urdu, Arabic	Yes	DT, CatBoost, XGB, LR, BiLSTM, CNN, BGRU, BERT, DistilBERT	Yes

[1] Prohibited in Russian Federation.

platform. Data collection spanned from September 2023 to March 2024, offering a robust foundation for conducting in-depth analyses and investigations into the dynamics of hopeful communication within the digital landscape. After collecting the samples from Twitter, we combine our data from English, Arabic, and Urdu into a single CSV file. This combined dataset is called Posi-Vox-2024. «Posi Vox,» derived from «Posi» (positive) and «Vox» (voice), focuses on hope speech and aims to detect positive discourse across multilingual communities. Figure 1 shows the proposed methodology and design of the study, outlining the process of analyzing hope speech in mixed texts commonly found in social media discussions within multilingual communities. The term «multilingual hope speech detection» refers to this unified approach that processes and interprets mixed-language texts to enhance sentiment analysis across diverse online communities. Our proposed model captures linguistic nuances without translation, making it highly relevant for multilingual social media platforms. It offers a scalable solution for detecting hope speech in mixed-language content, providing greater flexibility and robustness compared to traditional monolingual models, thereby enhancing sentiment analysis and fostering positive discourse in diverse online communities.

Data Preprocessing

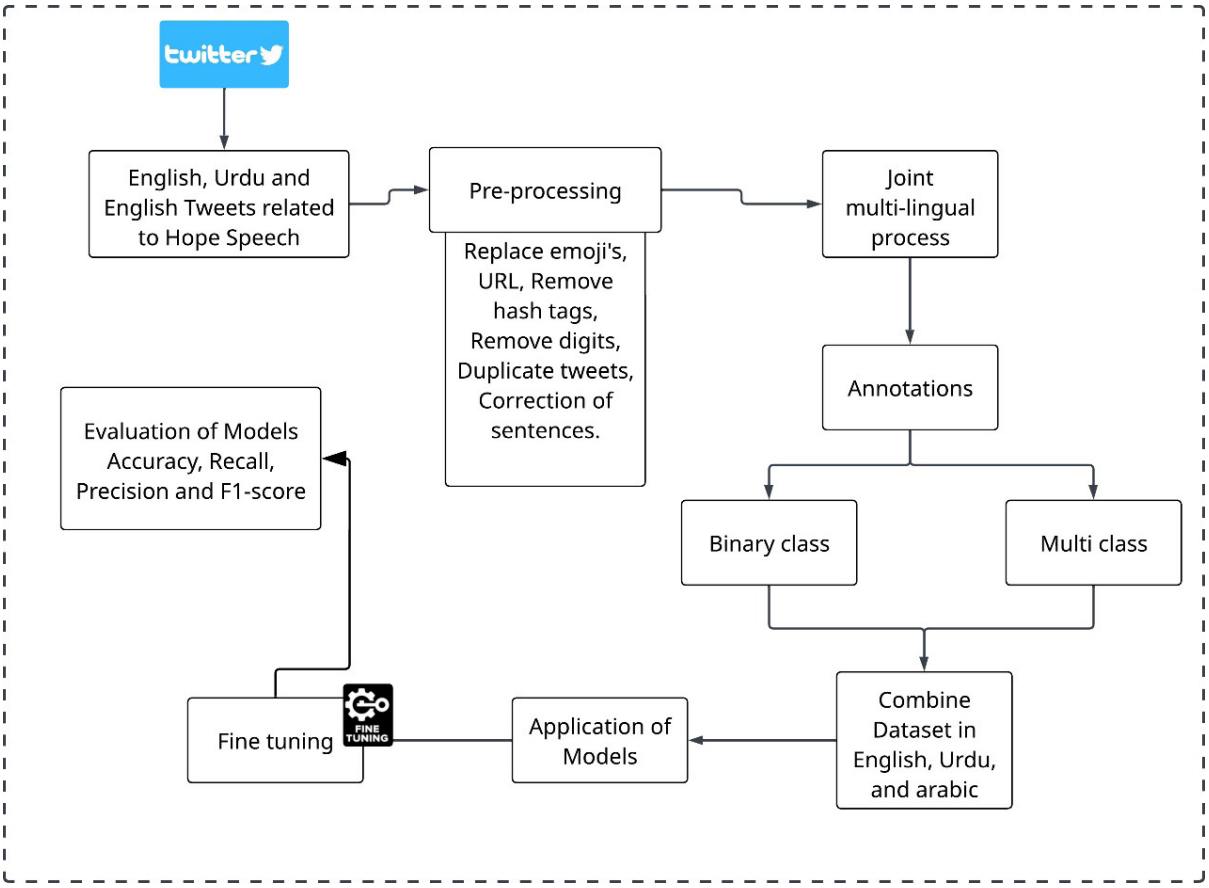
The Tweepy[2] API was developed for providing functionalities to filter tweets based on different criteria, such as date, location, language, and tweet id. Specifically, we utilized the date and language attributes to scrape tweets in the English, Urdu, and Arabic languages. Due to the extensive noise present in social media's textual content, we conducted various preprocessing procedures:

1. Eliminating URLs, user mentions in the form @use, and HTML Tags.
2. Removal of punctuation marks from the text.
3. Remove duplicates and less than 20-character tweets.
4. Uppercase Text is transformed to lowercase.
5. Replace the emoji with a corresponding text; as we know, emoji's play an essential role in detecting tweets.
6. Removal of the Digits in the tweets.
7. Decodes all the short text such as thnx to Thanks, plz to please, etc.

After processing 80,000 tweets, only 18,362 original tweets remained in Urdu, English, and Arabic to create a joint multilingual dataset.

Figure 1

Proposed Methodology and Design



[2] https://www.tweepy.org/ Last visited: 11-10-2024.

Annotation Process

Annotation Guidelines

Based on the definition of hope provided by psychologists, we categorized the tweets into two classes. The primary class encompassed tweets expressing hope, whereas the secondary class comprised tweets devoid of any sense of hopeful-ness. This classification methodology enables us to analyze and interpret the presence or absence of hope within tweet content, allowing for deeper insights into user sentiments and emotional expressions on social media platforms. In the next phase of analysis, we categorized tweets into various types of hope by examining the specific features and characteristics present in the content. We implemented specific guidelines for the primary and secondary categorization of tweets, which are detailed along with the examples in Tables 2 and 3.

1. NHS: The tweet does not convey any sense of hope, aspiration, desire, or anticipation of the future.
2. Generalized Hope: This form of hope is characterized by a general sense of optimism and hopefulness that is not tied to any particular event or outcome.
3. Unrealistic Hope: often manifests as a wish for something to materialize despite its likelihood of being remote or virtually non-existent. Occasionally, individuals may harbor hope for irrational events or outcomes stemming from emotions such as anger, sadness, or depression.
4. Realistic Hope: This type of hope entails anticipating something that is reasonable, meaningful, and within the realm of possibility. There is a strong likelihood that anticipated events or outcomes will occur.

Annotator Selection

We explicitly avoided selecting annotators for the Posi-Vox-2024 dataset based on racial information, thereby demonstrating our unwavering dedication to promoting a culture of equity and diversity, while upholding the integrity of the dataset. We made a deliberate effort to record the nationality of annotators while avoiding the consideration of racial information. This approach allowed us to monitor the geographical diversity of our annotators in an unbiased manner, as shown in Table 4, without incorporating any biases related to racial characteristic.

Table 2

Binary Class Hope Speech

No.	Tweets	Category
1	جب کسی کام کا ارادہ کر لیں تو مکمل اللہ پہ بھروسہ کریں اور ہر شک و شبہ کو دل سے نکال دیں کیونکہ جس کے سپرد آپ نے اپنے معاملات کیے ہیں وہ بہترین کارساز ہے (When you intend to do something, put your full trust in Allah and remove all doubts from your heart because He is the best doer to whom you have entrusted your affairs.)	Hope
2	Nobody cares, you are undesired, and these no standard black men are hyping u up.	Not hope

Table 3

Multi-Class Hope Speech

No.	Tweets	Category
1	Embracing each day with optimism, believing in brighter tomorrows, and trusting in the journey ahead.	Generalized Hope
2	Have faith that your life will improve and that everything will work out for the best. Trust that your health will improve and love will come your way.	Realistic Hope
3	I dream of flying without wings, soaring above the clouds, defying gravity's hold.	Unrealistic Hope

Table 4

Annotators Based On Geographical Area

Language	Country	Male	Female	Undergraduate	Postgraduate
English	UK	2	0	2	0
Urdu	Pakistan	2	1	2	1
Arabic	UAE	2	0	2	0

Annotation Procedure

Selected annotators were provided with comprehensive annotation guidelines and sample annotations in the 'Annotation Setup' section. All annotators listed in Table 4 possess strong annotation skills, holding both undergraduate and postgraduate degrees, coupled with experience in NLP, machine learning, and deep learning. To supervise the annotation process, individual Google Forms were created for each annotator, and weekly meetings were scheduled to assess the progress of the annotation and identify any challenges encountered during the process. Figure 2 illustrates the steps involved in corpus creation for hope speech detection from social media tweets. Initially, the dataset undergoes binary classification to distinguish tweets exhibiting signs of hope from those that do not. Subsequently, within the affirmative class, further classification identifies specific emotional categories such as generalized hope, realistic hope, and unrealistic hope.

Dataset Statistic

Figure 3 depicts a word cloud comprising keywords extracted from tweets in a multilingual dataset related to the topic of hope speech. Figure 4 depicts the distribution of labels for both the binary and multiclass classifications. We collected equal data related to hope and not-hope class categories to show the data balance, and we needed to further categorize hope categories into multiple classes, such as generalized hope, realistic hope, and unrealistic hope based on emotions.

The Key characteristics of the hope speech dataset include total tweets (n=18362), total vocabulary size (n=105777), the total number of words (n=499486), the total number of characters (n=2583769), the average number of words (n=27.32), and the average number of character (n=141.56), as outlined in Table 5.

Data Augmentation

In order to improve the performance and robustness of our proposed model, we employed a data augmentation back translation technique. We utilized Google Translate API for the back translation process due to its wide coverage of languages and high translation quality. Custom scripts were developed to automate the translation process and handle large volumes of text efficiently. After back translation, we performed a manual quality check on a sample of the augmented data to ensure that the meaning of the original text was preserved and that no significant loss of information occurred during translation.

Ethical Concern

Figure 2

Annotation Procedure of Hope Speech Detection

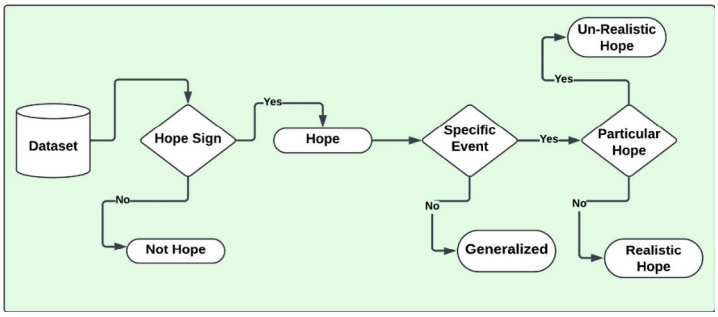


Figure 3

Word Cloud of Hope Speech Dataset

Figure 4

Label Distribution of Binary and Multi class in the entire Dataset

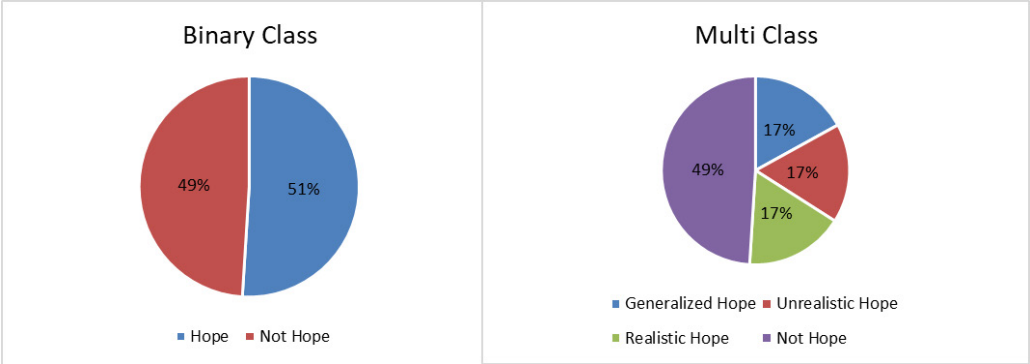


Table 5

Statistics of the Dataset

Class	Tweets	Words	Avg. Words	Characters Avg.	Characters	Vocabulary
Generalized hope	3082	82046	26.61	139.39	429603	19617
Realistic hope	3156	90355	28.63	150.12	473791	21286
Unrealistic hope	3074	83248	27.08	137.90	423933	20693
Not hope	9050	243837	26.94	138.83	1256442	44181
Total	**18362**	**499486**	**27.32**	**141.56**	**2583769**	**105777**

Data collected from Twitter is highly sensitive, and we highlight the privacy measures implemented in the data-annotation processes. The identities of the involved individuals remained hidden, and our annotator identified a name associated with a politician or celebrity. They adhered to a strict protocol of non-engagement, refraining from attempting to establish contact with such individuals.

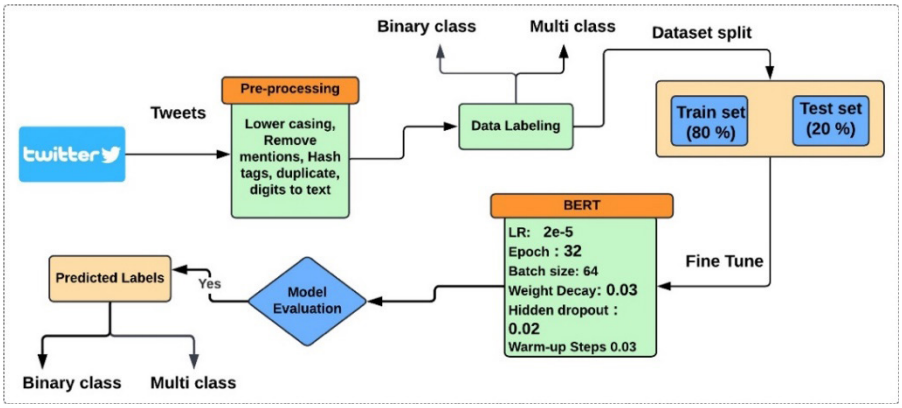
Methods for Hope Speech Detection

To demonstrate how our proposed Posi-Vox-2024 corpus can be used to develop, evaluate and compare methods for hope speech detection task, we applied and compared state-of-the-art four machine learning models: (i) Decision Tree (DT), (ii) CatBoost (CB), (iii) Extreme Gradient boosting (XGB), and (iv) Logistic Regression (LR); three Deep learning models: (i) Bidirectional Long Short-Term Memory (BiLSTM), (ii) Convolutional Neural Network (CNN), and (iii) Bidirectional Gated Recurrent Unit (BGRU); and two transfer learning models: (i) pre-train BERT, and (ii) distilBERT.

Our best performing model is based on the BERT architecture, leveraging transformer layers to capture contextual relationships in multilingual text. After preprocessing with appropriate tokenizers for English, Urdu, and Arabic, we fine-tuned a pre-trained BERT model on our annotated da-

Figure 5

BERT-Based Model Training Pipeline for Multilingual Text Classification



taset using cross-entropy loss and an Adam optimizer with a learning rate of 2e-5. The dataset was partitioned into 80% for training and 20% for testing. To ensure reproducibility, configurations, including batch size, number of epochs, and evaluation metrics, along with optimum hyper-parameter values, are presented in Table 9. A diagram illustrating the BERT architecture and data flow is provided in Figure 5 to clarify how the model processes multilingual input and predicts hope speech.

RESULTS

In this section, we present the results of various machine learning, deep learning, and transformer-based models applied to the task of multilingual hope speech detection. These models were evaluated on both binary and multi-class classification tasks using our proposed Posi-Vox-2024 corpus, which includes English, Urdu, and Arabic text. Tables 6, 7, 8, 10, and 11 present the Precision, Recall, F-1 score, and Accuracy results obtained by applying state-of-the-art machine learning algorithms such as Decision Tree (DT), Categorical Boosting (CatBoost), Extreme Gradient Boosting (XGB), and Logistic Regression (LR). For deep learning, we utilized Convolutional Neural Network (CNN), Bidirectional Gated Recurrent Unit (BGRU), and Bidirectional Long Short-Term Memory (BiLSTM) models. In the transformer category, we employed Bidirectional Encoder Representations from Transformers (BERT) and Distilled BERT (DistilBERT) on our proposed Posi-Vox-2024 corpus. Our experiments focused on identifying the most suitable model for handling hope speech across languages, systematically tuning hyperparameters for each model and analyzing their performance based on metrics such as accuracy, precision, recall, and F1-score. The following subsections provide detailed results for each category of models.

Machine Learning

Table 6 shows the results attained by the various machine learning models using TF-IDF word embedding for hope speech detection, classified into binary and multi-class tasks. For binary class of hope speech detection, the DT, CatBoost, XGB, and LR models show F1-scores ranging from 0.70 to 0.73, with LR achieving the highest precision, recall, F1-score, and accuracy of 0.75. In the multi-classification hope speech detection task, again LR performed better than all other models, achieving F1-score of 0.61. CatBoost and XGB shows competitive performance with Accuracy rates of 0.58, hence LR outperforms the other models in both binary and multi-class tasks, achieving the highest Precision, recall, F1-score, and accuracy.

Table 6

Results of Machine Learning Models

Class	Model	Precision	Recall	F1-score	Accuracy
Binary class	DT	0.72	0.72	0.72	0.72
	Catboost	0.73	0.73	0.73	0.73
	XGB	0.72	0.71	0.7	0.71
	LR	0.75	0.75	0.75	0.75
Multi class	DT	0.59	0.59	0.59	0.59
	Catboost	0.57	0.58	0.53	0.58
	XGB	0.57	0.58	0.54	0.58
	LR	0.61	0.61	0.61	0.61

Deep Learning

Table 7 presents the performance metrics for three deep learning models such as CNN, BGRU, and BiLSTM on binary and multi-class classification tasks. For the binary classification task, all three models perform similarly, with the CNN and BiLSTM models achieving a precision, recall, and F1-score of 0.75, while the BGRU model has slightly lower values (0.74). The accuracy for all three models is also consistent, at 0.75 for CNN and BiLSTM, and 0.74 for BGRU. In the multi-class classification task, the models show a decrease in performance across all metrics. The CNN and BGRU models have precision, recall, F1-score, and accuracy around 0.56, while BiLSTM performs slightly better with 0.62 for all metrics. This suggests that while the models perform well on binary classification, they face more challenges with multi-class classification.

Table 7

Results of Deep Learning Models

Class	Model	Precision	Recall	F1-score	Accuracy
Binary class	CNN	0.75	0.75	0.74	0.75
	BGRU	0.74	0.74	0.74	0.74
	BiLSTM	0.75	0.75	0.75	0.75
Multi class	CNN	0.56	0.56	0.56	0.56
	BGRU	0.55	0.55	0.55	0.55
	BiLSTM	0.62	0.62	0.62	0.62

Transformer Results

The table 8 summarizes the performance metrics Precision, Recall, F1-score, and Accuracy—for binary and multi-class classification tasks. In binary classification, BERT achieves higher precision, recall, F1-score, and accuracy (all at 0.78) compared to DistilBERT, which has slightly lower scores (F1-score of 0.75 and accuracy of 0.76). In multi-class classification, both models show a performance drop; BERT attains an F1-score of 0.65 and accuracy of 0.66, while DistilBERT has a slightly lower F1-score and accuracy of 0.64. Overall, BERT outperforms DistilBERT across both tasks.

Table 8

Transformer Results

Class	Model	Precision	Recall	F1-score	Accuracy
Binary class	Bert	0.78	0.78	0.78	0.78
	DistilBert	0.76	0.76	0.75	0.76
Multi class	Bert	0.66	0.66	0.65	0.66
	DistilBert	0.64	0.64	0.64	0.64

Table 9 presents the optimal fine-tuning parameters for a pre-trained BERT model for both binary and multi-class classification tasks. The best hyper-parameters were recognized through grid search, considering the following ranges: learning rates are 1e-5, 1e-2, 2e-5, 3e-5, 3e-4, epochs are 9, 32, 64, batch sizes are 64, 128, 512, weight decay values from 0.01 to 0.1, hidden dropout rates of 0.02 and 0.1, and warm-up steps from 0.03 to 0.1. These settings ensure balanced training efficiency and robust model performance across various classification problems.

Table 9

Optimum Values Identified for the Hyper-Parameters of the Bert Model

Hyper-parameter	Grid search
Learning rate	1e-5,1e-2, 2e-5, 3e-5, 3e-4
Epoch	3, 9, 32
Batch size	32, 64, 128
Weight Decay	0.01–0.1
Hidden dropout	0.02, 0.1
Warm-up Steps	0.03–0.1

Error Analysis

Table 10 shows class-wise scores, while Figure 6 shows the confusion matrix for both binary and multiclass classification in percentage achieved by our proposed model. Notably, our model demonstrated better performance in the not hope class in terms of precision. In classifying Unrealistic hope, our proposed model performs better than all other labels of hope class while showing comparatively lower accuracy in distinguishing between generalized and realistic hope categories.

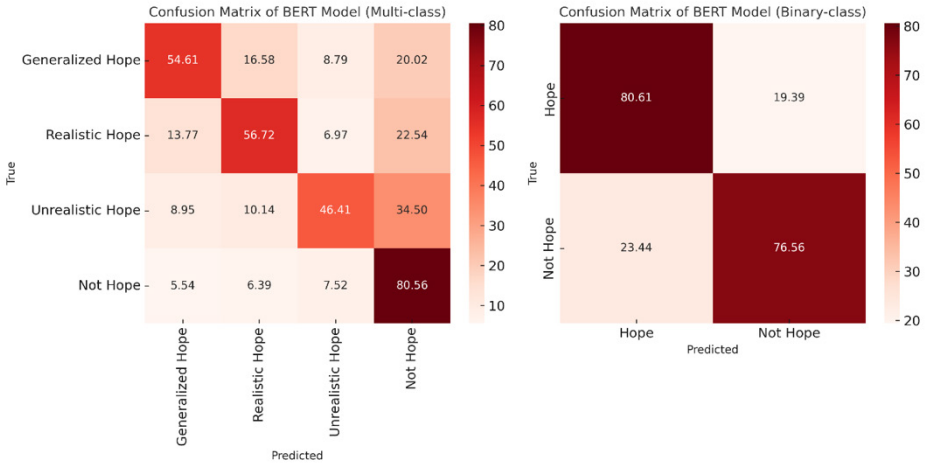
Table 10

Class Wise Score for the Proposed Methodology

Class	Categories	Precision	Recall	F1-Score	Support	Accuracy
Binary class	Hope	0.78	0.79	0.78	3666	0.78
	Not hope	0.79	0.78	0.78	3631	
Multi class	Generalized hope	0.50	0.63	0.55	1194	0.66
	Realistic hope	0.57	0.53	0.55	1220	
	Unrealistic hope	0.63	0.41	0.50	1252	
	Not hope	0.75	0.80	0.77	3631	

Figure 6

Confusion Matrix of Proposed Methodology



The pre-trained BERT model demonstrated a notable performance gain over traditional machine learning models, with a binary classification accuracy of 0.78 compared to traditional machine learning, such as LR 0.75, resulting in a performance improvement of approximately 4%. For multi-class tasks, BERT achieved 0.66, outperforming the LR 0.61, indicating a performance improvement of about 8.20%. This suggests BERT's superior contextual understanding and handling of nuanced language, especially in a multilingual setting. Thus, BERT's advanced language modeling provides significant advantages in detecting hope speech across different languages. Table 11 shows the outcomes of the top performing models of each learning approach in binary and multi class.

Table 11

Top Performing Models in Each Learning Approach

Class	Model	Learning approach	Accuracy
Binary class	LR	Machine learning	0.75
	BiLSTM	Deep learning	0.75
	BERT	**Transformer**	**0.78**
Multi class	LR	Machine learning	0.61
	BiLSTM	Deep learning	0.62
	BERT	**Transformer**	**0.66**

DISCUSSION

This study explores a valuable set of features that effectively detect hope speech expressions in Twitter tweets. This research offers meaningful insights for online users and society, aiming to foster peace and positivity. Hope is often linked to providing encouragement, support, reassurance, suggestions, or inspiration to individuals during times of illness, stress, loneliness, or depression (Snyder et al., 2002). The literature review has primarily focused only binary class approaches to identifying hope speech detection on different social media platforms, with limited work on developing a multilingual framework. To address this research gap, we built a multilingual tool that combines joint multilingual methodology to tackle the task of hope speech detection in English, Urdu and Arabic languages. Our proposed tool was trained and tested on a multilingual dataset to uncover practical insights and ensure its applicability on real time. Our findings show that our proposed method is very effective and power tool to identify the hope speech detection in Twitter post. We utilized the power of transfer learning methods by fine-tuning the pre-trained BERT model and added a fresh contribution that attained 78% accuracy in binary class and 66% in multi classification. Furthermore, our proposed methodology outperformed the four baselines such as LR, XGB, CB and DT. Thus, based on these results, our proposed framework can be employed for other Multi-lingual Text Classification problems in similar fields.

There are several limitations in this study. Firstly, collecting and annotating hope speech data in English, Urdu, and Arabic pose several challenges. One of the main difficulties is identifying native speakers fluent in these languages who also possess knowledge in NLP and machine learning for accurate and reliable annotations. Secondly, during annotation process, we encountered numerous tweets that expressed hope but had a negative undertone. For instance, in Urdu, tweets such as " میری امید ہے کہ میرے دشمنوں کو تباہی کا سامنا کرنا پڑے، ان کی تباہی میری خوشی ہوگی #USER-" ("My hope is that my enemies should suffer destruction, their destruction will be my joy. #USER») present a challenge. Although this tweet conveys hope, its primary sentiment is negative, further complicating the annotation process. Thirdly, Urdu and Arabic are considered low-resource languages in the context of machine learning and deep learning, making it more challenging to understand and process them, which in turn hinders the development of robust models for detecting hope speech. Fourthly, despite its notable performance in multilingual hope speech detection, the proposed work has limitations. The Posi-Vox-2024 dataset is limited in size and diversity, thereby affecting its generalizability. Language-specific nuances and code-switching have not been fully addressed, potentially impacting the classification accuracy. The model's complexity and resource requirements limit accessibility, and its performance may degrade over time owing to the dynamic nature of the social media discourse.

CONCLUSION

Social media has become a powerful space for public dialogue, influencing opinions and the emotional landscape of communities. Until now, most research has focused on addressing negativity in the English language, particularly hate speech detection. This study highlights the critical need for multilingual hope speech detection (MHSD) in social media discourse, particularly focusing on the Urdu and Arabic languages, which has been overlooked in existing research. To achieve this objective, we address the two level text classification and built a comprehensive dataset named as Posi-Vox-2024 based on three languages such as English, Urdu and Arabic to tackle the challenges of multilingualism and improve communication across different backgrounds. By creating a multilingual dataset and employing state-of-the-art transfer learning models with fine-tuning, we effectively addressed the challenges associated with identifying hope speech in English, Arabic and Urdu. The results indicate that our proposed framework, utilizing pre-trained BERT model, significantly outperformed four baseline models (DT, XGB, Catboost, and LR), achieving accuracies of 0.78 in binary class and 0.66 in multi class. These findings underscore the importance of promoting positive discourse online and demonstrate the potential of hope speech as a means to foster healthier and more constructive interactions within communities. Further exploration could focus on expanding the dataset and incorporating additional languages to enhance the generalizability and robustness of the proposed framework.

ACKNOWLEDGMENTS

The work was done with partial support from the Mexican Government through the grant A1-S-47854 of CONAHCYT, Mexico, grants 20241816, 20241819, 20240936 and 20240951 of the Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico. The authors thank the CONAHCYT 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.

DATA AVAILABILITY

Data will be made available on request.

DECLARATION OF COMPETITING INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Muhammad Ahmad: Conceptualization; Data curation; Methodology; Resources; Software; Visualization; Writing – original draft; Writing – review & editing.

Usman Sardar: Data curation; Formal analysis; Methodology.

Humaira Farid: Investigation; Visualization; Writing – review & editing.

Iqra Ameer: Data curation; Formal analysis; Methodology; Software; Writing – original draft; Writing – review & editing.

Muhammad Muzamil: Data curation; Methodology; Software.

Ameer Hmaza: Data curation; Methodology.

Grigori Sidorov: Conceptualization; Resources; Supervision; Validation.

Ildar Batyrshin: Conceptualization; Project administration; Resources; Supervision; Validation; Writing – original draft; Writing – review & editing.

REFERENCES

Alawadh, H. M., Alabrah, A., Meraj, T., & Rauf, H. T. (2023). English language learning via YouTube: An NLP-based analysis of users' comments. *Computers*, *12*(2), 24.

Anand, M., Sahay, K. B., Ahmed, M. A., Sultan, D., Chandan, R. R., & Singh, B. (2023). Deep learning and natural language processing in computation for offensive language detection in online social networks by feature selection and ensemble classification techniques. *Theoretical Computer Science, 943*, 203-218.

Anjum, & Katarya, R. (2024). Hate speech, toxicity detection in online social media: a recent survey of state of the art and opportunities. *International Journal of Information Security, 23*(1), 577-608.

Arif, M., Shahiki Tash, M., Jamshidi, A., Ullah, F., Ameer, I., Kalita, J., ... & Balouchzahi, F. (2024). Analyzing hope speech from psycholinguistic and emotional perspectives. *Scientific Reports*, *14*(1), 23548.

Austin, D., Sanzgiri, A., Sankaran, K., Woodard, R., Lissack, A., & Seljan, S. (2020). Classifying sensitive content in online advertisements with deep learning. *International Journal of Data Science and Analytics, 10*(3), 265-276.

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

Chakravarthi, B. R. (2022). Hope speech detection in YouTube comments. *Social Network Analysis and Mining*, *12*(1), 75.

Chakravarthi, B. R. (2022). Multilingual hope speech detection in English and Dravidian languages. *International Journal of Data Science and Analytics, 14*(4), 389-406. https://doi.org/10.1007/s41060-022-00341-0

Chinnappa, D. (2021). Dhivya-hope-detection@ LT-EDI-EACL2021: multilingual hope speech detection for code-mixed and transliterated texts. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion* (pp. 73-78). Publisher. https://aclanthology.org/2021.ltedi-1.11

Davidson, T., Bhattacharya, D., & Weber, I. (2019). *Racial bias in hate speech and abusive language detection datasets*. arXiv preprint arXiv:1905.12516.

Gowen, K., Deschaine, M., Gruttadara, D., & Markey, D. (2012). Young adults with mental health conditions and social networking websites: seeking tools to build community. *Psychiatric Rehabilitation Journal, 35*(3), 245.

Ghanghor, N., Ponnusamy, R., Kumaresan, P. K., Priyadharshini, R., Thavareesan, S., & Chakravarthi, B. R. (April). IIITK@ LT-EDI-EACL2021: Hope speech detection for equality, diversity, and inclusion in Tamil, Malayalam and English. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion* (pp. 197-203). Publisher.

Irfan, A., Azeem, D., Narejo, S., & Kumar, N. (2024, January). Multi-Modal Hate Speech Recognition Through Machine Learning. In *2024 IEEE 1st Karachi Section Humanitarian Technology Conference (KHI-HTC)* (pp. 1-6). IEEE.

Khouloud Mnassri, Reza Farahbakhsh, Razieh Chalehchaleh, Praboda Rajapaksha, Amir Reza Jafari, Li, G., & Crespi, N. (2024). A survey on multi-lingual offensive language detection. *PeerJ. Computer Science*, *10*, e1934–e1934. https://doi.org/10.7717/peerj-cs.1934

Kogilavani, S. V., Malliga, S., Jaiabinaya, K. R., Malini, M., & Kokila, M. M. (2023). Characterization and mechanical properties of offensive language taxonomy and detection techniques. *Materials Today: Proceedings, 81*, 630-633. https://doi.org/10.1016/j.matpr.2021.04.102

Kumar, A. Saumya, S., & Roy, P. (2022, May). SOA_NLP@ LT-EDI-ACL2022: An ensemble model for hope speech detection from YouTube comments. In *Proceedings of the second workshop on language technology for equality, diversity and inclusion* (pp. 223-228). DOI: 10.18653/v1/2022.ltedi-1.31

Lee, Y., Yoon, S., & Jung, K. (2018). *Comparative studies of detecting abusive language on twitter*. arXiv preprint arXiv:1808.10245.

Louati, A., Louati, H., Albanyan, A., Lahyani, R., Kariri, E., & Alabduljabbar, A. (2024). Harnessing machine learning to unveil emotional responses to hateful content on social media. *Computers, 13*(5), 114.

Malik, M. S. I., Nazarova, A., Jamjoom, M. M., & Ignatov, D. I. (2023). Multilingual hope speech detection: A Robust framework using transfer learning of fine-tuning RoBERTa model. *Journal of King Saud University-Computer and Information Sciences, 35*(8), 101736.

Mnassri, K., Farahbakhsh, R., Chalehchaleh, R., Rajapaksha, P., Jafari, A. R., Li, G., & Crespi, N. (2024). A survey on multi-lingual offensive language detection. *PeerJ Computer Science, 10*, e1934.

Nagar, S., Barbhuiya, F. A., & Dey, K. (2023). Towards more robust hate speech detection: using social context and user data. *Social Network Analysis and Mining, 13*(1), 47.

Nath, T., Singh, V. K., & Gupta, V. (2023). *BongHope: An annotated corpus for Bengali hope speech detection*. Research Square. https://doi.org/10.21203/rs.3.rs-2819284/v1

Palakodety, S., KhudaBukhsh, A. R., & Carbonell, J. G. (2020). Hope speech detection: A computational analysis of the voice of peace. In *ECAI 2020* (pp. 1881-1889). IOS Press.

RamakrishnaIyer LekshmiAmmal, H., Ravikiran, M., Nisha, G., Balamuralidhar, N., Madhusoodanan, A., Kumar Madasamy, A., & Chakravarthi, B. R. (2023). Overlapping word removal is all you need: Revisiting data imbalance in hope speech detection. *Journal of Experimental & Theoretical Artificial Intelligence*, *36*(8), 1837–1859. https://doi.org/10.1080/0952813X.2023.2166130

Roy, P., Bhawal, S., Kumar, A., & Chakravarthi, B. R. (2022, May). IIITSurat@ LT-EDI-ACL2022: Hope speech detection using machine learning. In *Proceedings of the Second Workshop on Language* Technology for Equality, Diversity and Inclusion (pp. 120-126). Publisher. https://aclanthology.org/2022.ltedi-1.13

Schmidt, A., & Wiegand, M. (April). A survey on hate speech detection using natural language processing. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media* (pp. 1-10). Publisher.

Snyder, C. R., Rand, K. L., & Sigmon, D. R. (2002). Hope Theory: A Member of the Positive Psychology Family. In C. R. Snyder, & S. J. Lopez (Eds.), *Handbook of Positive Psychology* (pp. 257-276). Oxford University Press.

Subramanian, M., Sathiskumar, V. E., Deepalakshmi, G., Cho, J., & Manikandan, G. (2023). A survey on hate speech detection and sentiment analysis using machine learning and deep learning models. *Alexandria Engineering Journal, 80*, 110-121.

Wang, Z., & Jurgens, D. (2018). It's going to be okay: Measuring access to support in online communities. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 33-45). Publisher.

Yates, A., Cohan, A., & Goharian, N. (2017). *Depression and self-harm risk assessment in online forums*. arXiv preprint arXiv:1709.01848.

Yenala, H., Jhanwar, A., Chinnakotla, M. K., & Goyal, J. (2018). Deep learning for detecting inappropriate content in text. *International Journal of Data Science and Analytics, 6*, 273-286.

Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N., & Kumar, R. (2019). *Predicting the type and target of offensive posts in social media*. arXiv preprint arXiv:1902.09666.