RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt2541

Analysis of Emotion Detection from Code Mixed or Code-Switched Social Media Text using Deep Learning

Vinayak Malavade [a, b, *], Virat Giri [c]

[a] Department of Computer Science and Engineering, Sanjay Ghodawat University, Kolhapur, 416118, Maharashtra, India

[b] Department of Computer Engineering (Regional Language), Pimpri Chinchwad College of Engineering, Nigdi, Pune, 411044, Maharashtra, India

[c] Sanjay Ghodawat Institute, Kolhapur, 416118, Maharashtra, India

* Corresponding Author Email: vnmalavade@gmail.com

DOI: https://doi.org/10.54392/irjmt2541

Received: 06-12-2024; Revised: 18-06-2025; Accepted: 26-06-2025; Published: 01-07-2025

Abstract: Emotion detection plays a vital role in understanding user sentiment specially in this the era of digital communication. Due to exponential growth of internet and social media, social media became a huge source of information which gives insight in varieties of applications. Emotion detection is an emerging field of research. Users express their implicit feelings, thoughts on social media. Analyzing such data becomes challenging due to linguistic diversity, especially in code-mixed or code-switched content involving transliterated Hindi and English. This paper addresses the problem of emotion detection from such complex textual data. We have explored and compared performances of different deep learning models in handling code-mixed social media text. Our experiments demonstrate that among all tested architectures, the hybrid LSTM-CNN model shown the highest mean test accuracy of 79.60% and 0.4216 F1-score without data balancing. For balanced data CNN gives the highest accuracy of 77.79%, while the Bi-LSTM model gives the highest F1-score of 0.4978. This research demonstrates the effectiveness of deep learning for emotion detection in transliterated Hindi-English social media posts.

Keywords: Emotion Detection, Code Mixed Text, Code Switched Text, Social Media Text, Deep Learning

1. Introduction

Approximately 50.64% of the global population make use of internet and Social Media (SM) networks [1]. In last two decades the use of SM increased tremendously which results in huge amount of data is getting generated. Social networks became platform for internet users to express and share the thoughts [2]. In this era of evolution code mixing is became most common and comfortable way of communication during expressing thoughts or feelings. SM data contains hidden, implicit contextual information and extraction of such implicit information plays important role in different applications like artificial intelligence, marketing, etc. Sentiment Analysis (SA) is analysis of human language by extraction of feelings, thoughts etc. expressed in a piece of text, image, audio etc. and classifies as neutral, positive or negative. Emotion Detection (ED) is subtype of sentiment analysis which extracts fine grained emotions, feelings, thoughts of particular individual or user [3]. Emotion models mainly categories as dimensional and categorical. Dimensional emotions are continuous e.g. valence where, categorical emotions are discrete, such as anger, happiness, etc. [4]. Paul Ekman focused six emotions as Surprise, Sadness, Fear, Disgust, Anger and Happiness [5]. Two more emotions added such as trust and anticipation [6]. Figure 1. Gives the Paul Ekman's basic categories of emotions.

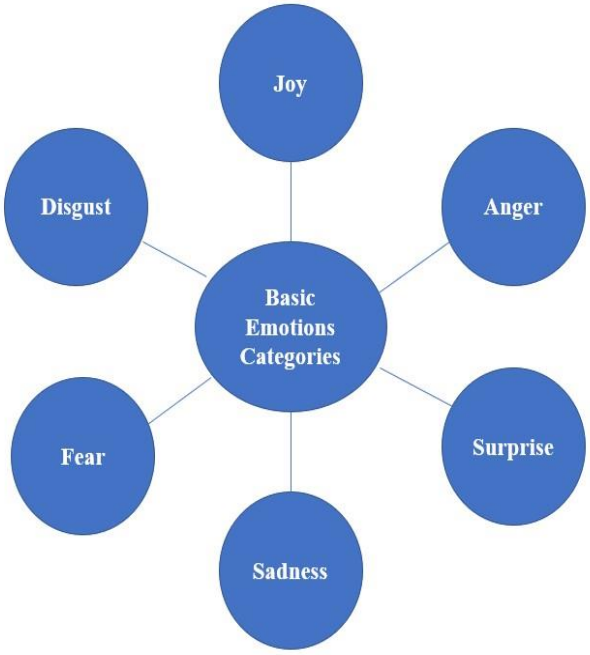


Figure 1. Paul Ekman's Basic Emotions

In general individuals' voices, words, expressions all represents their mood and feelings [7]. Psychosocial interventions can enhance ED and help reveal motivations for criminal actions [8]. ED and SA can be applied across different sectors to address challenges in analyzing human behavior [9]. SM posts are broadly seen and rich in emotional content [10]. Emotion analysis in language provides measurable insights into subjective expression, linking psychology, cognition, and linguistics [11].

ED plays a major role in applications like Healthcare, Market Research, Product Development etc. However, extracting emotions from social media text involves several challenges, especially when dealing with the linguistic diversity and informal nature of online communication which show different opportunities in field of Natural Language Processing (NLP). Major challenges in ED research is occurrence of Code-Mixed (CM) and Code-Switched (CS) language on social media platforms, where users mix multiple languages in the same sentence. This leads to issues like non-standard spellings, grammar inconsistencies, and variable word order. As per Census 2011 approximately 255 million peoples are proficient in two languages and 87 million peoples are trilingual in India. Thus, code-mixing is becoming most common way of communication [12]. In population 1.3 billion peoples, approximately 1600 different languages were spoken in India in which Hindi is widely used language. People have habit to use Hindi-English code-mixed (CM) language on SM [13]. Approximately 187 million users worldwide are active daily on tweeter [14]. Over 62 million tweets were used for automatic language detection on tweeter in which only half of the tweets were in English and others were mixed languages without including word-level CM [15].

India officially recognizes 22 languages, yet hundreds of regional languages and dialects are spoken in daily life [16]. CM refers to the integration of words from one language into another without altering the overall context. In contrast, code-switching (CS), also known as language alternation, involves switching between two or more languages within the same conversation or communicative setting. A deep understanding of CM and CS data is essential for enhancing communication, conducting accurate SA, and making communication effective for diverse languages [17]. These multilingual and informal communication patterns pose a major gap in existing Emotion Detection models, which are predominantly trained and tested on monolingual and well-formed texts. Most of traditional Deep Learning (DL) models are not completely adapted to handle the complexities of CM or CS data. CM increase complexity in sentiment and emotion analysis. SM posts, especially on platforms like Twitter, frequently merge languages, posing challenges in decoding the emotional and contexual content [18]. The absence of facial expressions and voice tone makes emotion analysis from text difficult [19]. There are various challenges in the open-source resources in Hindi language like imbalanced datasets and lack of understanding of textual nuances specific to Hindi [20]. Most studies on SA and ED are carried out in single language at single time, multilingual ED and CM environments are not fully explored [21]. Hindi-English CM language, referred as Hinglish, is a language in which speakers mix Hindi and English in conversations [22]. With the increasing occurrence of CM languages, there is need to analyze and understand this linguistic data. Language models developed for single languages like English or Hindi are quite robust and effective but they struggle to perform well with CM languages [23]. The quality of machine translation systems is further improved with the evolution of encoder-decoder architecture. The attention mechanism improves the accuracy of long sentence translation [24]. In multilingual societies, such as India, a large portion of social media content is code-mixed, where users blend languages like Hindi and English within the same text, resulting in what is commonly known as Hinglish [25]. Detection of emotion, and emotion intensity of memes is performed which shows new area of research [26].

The key focus of this work is to evaluate and compare performances of various DL models for ED from CM or CS social media text. In last two decades DL shown enhanced results in various field of applications. DL techniques like Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), attention mechanisms and transformers have shown significant success in many NLP tasks, their application to CM or CS emotion detection is still underexplored. In this study, we have built hybrid models by combining benefit of CNN and RNN and fine tune the model hyperparameters such as filter size, kernel size, number of layers etc. to detect emotions from CM or CS texts. To address lexical challenges, we construct two transliteration dictionaries containing 253,916 entries. One maps words from transliterated Hindi-English (Hinglish) to Devanagari Hindi words, and the other maps Devanagari Hindi to English words. In proposed hybrid modes CNN is used to extract contextual features and RNN is used to get sequential dependencies.

This paper is organized as Section 1 gives introduction. Literature review presented in section 2. Outline of work (methodology) explained in section 3. Performance analysis of proposed models described in section 4. Finally, section 5 discussed conclusion of this study.

2. Literature Survey

Need of ED in various fields like E-commerce, mental health etc. explore various approaches of ED from text, image etc. Lots of studies are done in the field of ED but still performances of available models of Machine Learning (ML), DL, rule based approach, hybrid

models etc. are lagging due different issues. This literature survey provides an overview of different approaches used for ED. Textual ED model were proposed using NLP and Neural Network where NN gives promising results [7]. Semantic analysis on text data is performed with the help of DL using Big data [8]. Emotion classifier is formed with the help of two million tweets using Transfer Learning and Pre-training model. The performance of RNN, LSTM, Bi-LSTM and GRU is evaluated using the ISEAR dataset where the result insights GRU outperforms with 60.26% accuracy, Bi-LSTM reached 59.30%, and LSTM reached 57.65 % accuracy [9]. Sparse and dense representations with ensemble approach are used using pre-trained, dense word embedding in different datasets [27]. Classification based on DL approach is proposed also, comparison of different pre-trained word embedding models is performed [28]. Meta heuristic DL approach were proposed for ED from live SM data based on linguistics [10]. CNN with Sequence and word embedding model is proposed for ED. Attention mechanism is used to focus on word using CNN. Results shows good precision and accuracy for detection of emotion from text [29]. Now a day's multilingual society facing problems of various languages. Code-mixing is common problem in modern societies in which users switches their languages or use mixed languages within a conversation. Many of existing models of ED are worked on monolingual text. Study on basic approaches, generic process, challenges of SA from CM text were discussed, summarized also, performances of various CM based studies compared [30]. Hindi-English and Spanish-English datasets, with 20,000 and 19,000 examples, achieved F1 scores of 75.0 and 80.6, respectively for language detection of word and sentence polarity [31]. Emotion identification and analysis plays important role in getting trends, reviews, events and human behaviour even limitations of availability of Hindi-English CM text resources [32]. Study of bilingual or multilingual text analysis is still lagging due to varieties of issues like linguistic structure, lexicons, diversity of linguistic communities, idioms, sarcasm etc. LSTM proven better accuracy than pretrained model with effectiveness of language-specific stop words [11]. A DL approach for Hindi-English CM tweets using dual-language word vectors and transformer architectures shows improved accuracy of 71.43% [33]. The study focuses on ED from bilingual code mixed or CS transliterated Hindi- English SM text. A novel prompt-based learning approach is used for CM or CD text classification [17]. Transformer techniques for translation of CM languages in roman script to English is focused. Existing work perform better than multilingual translation [16]. Datasets, works, evaluations, etc. for SA on CM social media data are outline [34]. Sentiment polarity at sentence level for Indian CM languages e.g. Hindi-English is identified using ensemble voting classifier with other classifiers and linear SVM by character n-grams TF-IDF [35]. CNN, LSTM and Bidirectional LSTM models are used to analyse sentiments in Hinglish CM Data in which CNN shows the highest accuracy, 75.25% [36]. Various pretrained models on CM and non-CM data models are focused [37]. Novel approach (encoder architecture for monolingual and bilingual features) for improving code-switched ED from text were proposed using parallel translation [38]. To get cross linguistic knowledge multitask learning approach using techniques of adapter fusion above multilingual language model is proposed [39]. Data with 12000 Hindi-English CM texts were collected and labelled with emotions. Feature vectors are created with pretrained models (bilingual) and for classification models deep neural networks were used. CNN-Bi-LSTM model gives better performances [13]. Models for emojis prediction from CM (English-Hindi) sentences are proposed, also new dataset (SENTIMOJI) is focused [18]. Emotion classification using DL and transformer-based model are studied and analyse, also majority voting-based ensemble learning for performance improvement with SemEval 2019 Task 3 (SemEval) public dataset is focused, where ensemble model shows improved performances [19]. BERT and CNN combine model proves better results than baseline model using the SemEval and ISEAR datasets with 94.70 % and 75.80 % accuracy for respective datasets [40]. Large Language Models were focused for ED in Hindi text by exploring GPT-3.5-turbo interface using different types of learning and tuned pretrained models [20]. Novel multilimbed CM emotion dataset (EmoMix-3L) with 1,071 instances of Bangla-English-Hindi were introduced also, 100,000 synthetic data instances were created in which Multilingual Representations for Indian Languages (MuRIL) proves better results [21]. SpaCy model is trained for emotion classification in Hindi-English CM utterances using machine translation [22]. Review based on tasks of emotion analysis, emotion frameworks, emotion subjectivities NLP applications were performed along with its lacunas [41]. Various methods for recognition of accurate emotion and reasoning were developed for Hindi-English CM dialogues and English dialogues with F1-scores 0.78 and 0.79 for different tasks [42]. Emotions identification and finding causes behind emotion flips in monolingual English and bilingual Hindi-English CM dialogues were focused which attained F1-scores of 0.70, 0.79, and 0.76 for different tasks [43]. Emotion recognition model of Hinglish conversations were focused which explores the usage of different ML and DL pretrained models [23].

Existing study of ED from Hindi-English code-mixed test with respect to datasets used, proposed models, performance and future direction summarized in Table 1. Along with its future scope. Current studies on ED shows the hybrid DL and transformer-based models proven better accuracy on code-mixed datasets.

Even though many studies have worked on ED for CM or CD texts but their performance remains limited, particularly in handling bilingual or CM content.

Table1. Comparison of Key Studies on ED in CM or CS Text

References	Type of Language	Dataset	Techniques	Algorithms/ Models	Performances (Accuracy in % and F1 score)	Challenges / Limitation / Future scope
[9]	Monolingual	ISEAR	DL	LSTM, BiLSTM, and GRU	GRU 60.26%, BiLSTM 59.3%, LSTM 57.65%	Model's performance can be optimized by tuning the hyperparameters
[27]	Monolingual	SemEval	DL	EN-TFIDF	EN-TFIDF 0.525(F1 score)	Transfer learning for multidomain emotion detection can be used
[28]	Monolingual	ISEAR, SemEval 2018 task_1, and 2019 task 3	DL	FastTextEmb	SemEval 2018 task_1 FastTextEmb 83.6 %, SemEval 2018 task _3 FastTextEmb 90.6%	Focus on personalized AI systems and social robots to enhance emotional intelligence is needed
[32]	Monolingual	Online (28667 tweets with 6 emotions)	ML	SVM classifier with 10-fold cross-validation	SVM 58.2	Word level annotate corpus with part-of-speech tags can be used, data size can be increase
[35]	Monolingual	Own Dataset	ML	SVM, MLP, Bi-LSTM, Voting Classifier	(F1 scores) SVM 0.557, Voting Classifier 0.569, MLP 0.53, Bi-LSTM 0.54	Character embeddings along with the word embeddings can be focus
[20]	Monolingual	Own Dataset	Transformer based approach	BERT and mBERT	BERT (on Hinglish dataset): 71.43%	Linguistic Complexities and Multimodal Emotion Detection can be focused
[40]	Monolingual	Semeval 2019 task3 dataset and ISEAR	Transformer based approach	BERT-CNN	BERT-CNN 94.7% for semeval2019 task3 and 75.8% for ISEAR dataset.	Model performance can be improved with other pretrained models like XLNet, RoBERTa

[11]	Bi-Lingual (Code-Mixed)	SemEval -2024 as Task 10 Dialogue dataset	DL	NLP, LSTM, Dist-Bert	Word GloVe 35, Dist-Bert 30, LSTM Model 37.8	Data enhancement and use of clustering techniques can be used
[36]	Bi-lingual (Code-Mixed)	Online (28667 tweets with 6 emotions)	DL	CNN, LSTM, Bi-LSTM	CNN 75.25 %, LSTM 66.80, Bi-LSTM 71.27, Bi-LSTM (Attention) 73.25	Transformer models such as BERT, RoBERTa, ALBERT can be used
[13]	Bi-lingual (Code-Mixed)	Own Dataset (12000 tweets with 3 emotions)	DL	LSTM, Bi-LSTM, CNN and hybrid models	1D-CNN 78.13, LSTM 80.62, Bi-LSTM 81.07, CNN-LSTM 82.85, CNN-BiLSTM 83.21	Fine emotions can be focus in large Indian code-mixed corpus
[14]	Bi-Lingual (Code-Mixed)	Own datset on benchmark SentiMix	Transformer based approach	Transformer-based model, XLM-RoBERTa (XLMR)	Transfer Learning-XLMR: 66.03, F1score 64.47	Code-mixed texts from multilingual communities can be explore.
[21]	Bi-Lingual (Code-Mixed)	EmoMix-3L	Transformer based approach	DistilBERT, roBERTa, HindiBERT, HingBERT, XLM-R, mBERT	XLM-R 0.51 F1 Score, mBERT 0.49 F1 Score, MuRIL 0.67 F1 Score	Large language models can be explored to fine-tuning on codemixed datasets
[23]	Bi-Lingual (Code-Mixed)	SemEval 2024 Task 10 dataset	Transformer based approach	SVM, MNB, RF, Hing BERT, Hing mBERT, Hing RoBERTa	SVM 44%, MNB 40%, RF 43%, Hing BERT 45%, Hing mBERT 44%, Hing RoBERTa 47%	To enhance model efficiency with huge data and explore other emotions
[33]	Bi-Lingual (Code-Mixed)	Own Dataset	Transformer based approach	CNN, LSTM, Bi-LSTM, Bi-LSTM atten, BERT, RoBERTa, ALBERT	CNN 63.00, LSTM 64.59, Bi-LSTM 66.37, Bi-LSTM atten 68.29, BERT 71.44, ALBERT 66.22, RoBERTa 70.06	Language-specific transformer embeddings can be emphasized

To address this problem, proposed study employs special dictionaries along with a novel combined LSTM-CNN model to detect emotions in bilingual CM or CS Hindi-English texts.

3. Methodology

Dataset and methodology used for ED presented in this section along with experimental setup, model validation, and limitations.

3.1 Dataset

The process of ED starts with data collection. Dataset used to perform ED collected from online, publicly available datasets with 10614 annotated tweets with four categories of emotion such as joy (5874), anger (3382), sadness (1285), fear (73) annotated text. Each tweet is labeled with one of four emotion categories. Datasets consist of CM and CD text of roman transliterated Hindi and English, along with pure English and pure Hindi texts. Dataset collected from online sources such as GitHub, Kaggle, and other public datasets. Some

samples of Hindi-English CM annotated texts extracted from SM are given below in Table 2.

3.2 Workflow

Following Figure 2. Represents step used to perform ED from CM or CD Text.

3.2.1 Input

Annotated Code-mixed text data with four emotions categories joy, anger, sadness, fear is passed as input for ED.

3.2.2 Preprocessing

Pre-processing step is used to remove unwanted noise from text. It is performed in two steps. First step is used to filter noise in text performed before converting Transliterate Hindi-English text to Devanagari Hindi which includes lower case, stop word removal, tokenization, lemmatization etc. Second step is performed after translation of whole dataset from Devanagari Hindi to English. Once we received translated dataset, we refine such English translated dataset to remove non-contextual text.

3.2.3 Language identification and Translation

Language identification and translation performed based on dictionary-based approach. In our current approach, we utilize two separate dictionaries for transliteration and translation of code-mixed Hindi-English text. Two dictionaries (transliterated Hindi to Devanagari Hindi and Devanagari Hindi to English) with number of words is prepared based on online available platforms such as give hub, Kaggle, open AI tool etc. Keyword based language translation is performed in two steps. In first step all preprocessed words are mapped to transliterated Hindi to Devanagari Hindi dictionary, if words are matching respective word will be replace by corresponding Devanagari Hindi words. In second step Devanagari Hindi words are converted to roman English word using Devanagari Hindi to English words dictionary. The current study does not specifically address ambiguous terms or idiomatic expressions and has limitations in handling idiomatic phrases and words not included in the dictionary.

3.2.4 Vectorization

Pre-processed English text is converted to vectors with the help of vectorization techniques. In this step textual context is converted to numeric representations. Text samples were vectorized using word embedding layers with vocabulary size of 10000 words, embedding dimensions 16, maximum input sentence length considered as 50 and post padding technique is used. Vocabulary size and embedding dimensions chosen as per computational complexity after several trial runs.

Table 2. Annotated samples of transliterated Hindi-English Code-Mixed Data

Sr. No	Transliterated Code Mixed or Code-Switched Text	Emotion
1	“Ma’am aap ki smile ko copyright kar dijiye. It gives immense happiness seeing your smile”	Joy
2	“Jab job nahi de sakte tu program offer kiyun krwate ho”	Sadness

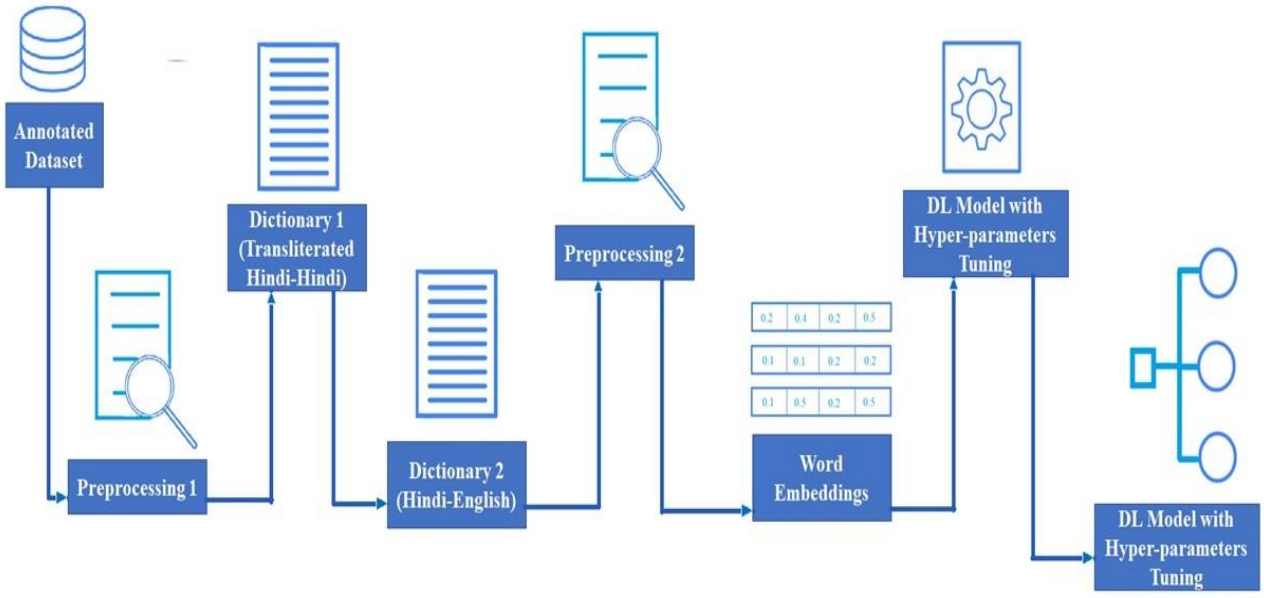


Figure 2. Proposed emotion detection framework

3.2.5 Model Building

Code-mixed and transliterated texts frequently include irregular grammatical structures, spelling mistakes etc. LSTM is suited for capturing contextual long-term dependencies in a text, whereas CNN effectively identifies emotion keywords which further used to detect patterns or features. Combining both feature extraction and sequential modelling provides a more robust framework for detecting emotions from noisy SM text. Hybrid Models of LSTM and CNN combine the benefits of both approaches. We have built various DL models such as CNN, LSTM, CNN-LSTM, CNN-LSTM-CNN, LSTM-CNN, and Bi-LSTM. To enhance generalization and reduce overfitting, models employed regularization techniques such as dropout layers (ranging from 0.3 to 0.5), batch normalization, and L2 kernel regularization. Hyperparameters including the number of convolutional filters, kernel sizes, LSTM units, learning rates, and dropout rates were systematically tuned through iterative experiments. The models are trained for multiple epochs with the help of early stopping and learning rate reduction callbacks to optimize training efficiency and to avoid overfitting. Data is spited into training, validation, and test using stratified sampling to preserve class distribution, ensuring robust evaluation. Multiple training runs with different random seeds conducted to verify stability and reproducibility of results.

3.2.6 Classification

For classification input texts first embedded into dense vectors and then processed by convolutional layers to extract features followed by LSTM layer for context. Final predictions made through dense layers with SoftMax activation. The models are trained using categorical cross-entropy and optimized with dropout and batch normalization to reduce overfitting.

3.3 Method Validation

All experiments were conducted using Python 3.8 and libraries including NLTK, TensorFlow, Keras. Dataset and dictionary words used to perform analysis of ED is prepared from online SM platforms. Various DL models are executed with input dataset. To overcome data imbalance in emotion classes, we have used oversampling technique for minority classes also used class weights during training. We evaluated several DL models under both balanced and imbalanced data conditions, using different metrics. To ensure statistical robustness, each model was trained and evaluated over five independent runs with different random seeds (e.g. 42,101) to account for variability due to initialization and data shuffling. A stratified split is used to divide the dataset into training, validation, and testing. Performance stability is assessed through multiple independent runs. For each model, we reported the average and standard deviation of test accuracy and macro F1-score as well as 95% confidence intervals, across all runs. To support our analysis, we also plotted the average confusion matrix for each model. Still models incorporate over-fitting as input dataset is unbalance which results in to overfitting.

3.4 Limitations

- Major limitation of this model is the availability of a limited vocabulary, i.e., dictionary size for Transliterated Hindi, Devanagari Hindi, and English Text.
- Size of the balanced annotated dataset affects the model performances. In some cases, the model gets overfitted.
- Use of a larger annotated dataset can enhance efficiency of models also help in minimizing overfitting problem.
- Use of non-dictionary words, misspelled words, shortcut words, and idioms also affect the model performances.

4. Results and Discussion

This section outlines the experimental findings of our emotion detection models applied to code-mixed social media texts. Various studies have explored emotion detection (ED) using diverse deep learning (DL) approaches, offering valuable insights into model effectiveness across different datasets. In the last decade, DL proven better results across many applications. Table 3. shows performances of various DL Models based on mean test accuracy and mean macro F1-score with and without data balancing. The Bi-LSTM model, without data balancing, gives a mean test accuracy of 0.7634 ± 0.0050 and a macro F1-score of 0.4209 ± 0.0074, but showed poor performance of minority class emotions, suggesting a strong skew toward the majority classes. When balancing was applied, test accuracy remained similar at 0.7614 ± 0.0070, while the macro F1-score improved noticeably to 0.4978 ± 0.0226. The CNN model had a higher test accuracy of 0.7952 ± 0.0043 and a similar macro F1-score of 0.4178 ± 0.0026 without balancing. After balancing, its accuracy slightly dropped to 0.7779 ± 0.0071, but the macro F1-score increased to 0.4604 ± 0.0115. The LSTM model gave results close to CNN, with a test accuracy of 0.7939 ± 0.0024 and macro F1-score of 0.4177 ± 0.0015 without balancing. However, when balancing was used, its performance dropped significantly, showing a test accuracy of 0.5133 ± 0.0862 and 0.1787 ± 0.0325 macro F1-score. The CNN-LSTM model shown a test accuracy of 0.7870 ± 0.0053 and a macro F1-score of 0.4137 ± 0.0025 without balancing. With data balancing, the accuracy declined to 0.7479 ± 0.0169, while the macro F1-score to 0.4776 ± 0.0096. Similarly, the CNN-LSTM-CNN architecture showed a

test accuracy of 0.7891 ± 0.0053 and a macro F1-score of 0.4148 ± 0.0029 without balancing. When balancing was applied, the accuracy slightly reduced to 0.7625 ± 0.0058, but the macro F1-score improved to 0.4745 ± 0.0078. The LSTM-CNN model gives the highest test accuracy of 0.7960 ± 0.0028 and 0.4216 ± 0.0040 macro F1-score without data balancing. After applying balancing, the test accuracy dropped to 0.7557 ± 0.0017, but the macro F1-score increased significantly to 0.4863 ± 0.0079. These observations highlight the trade-off between accuracy and balanced class performance, emphasizing the value of macro F1-score in evaluating models trained on imbalanced data. Figure 3. Represents the aggregated results and confusion matrix of performances of various DL models with balanced dataset.

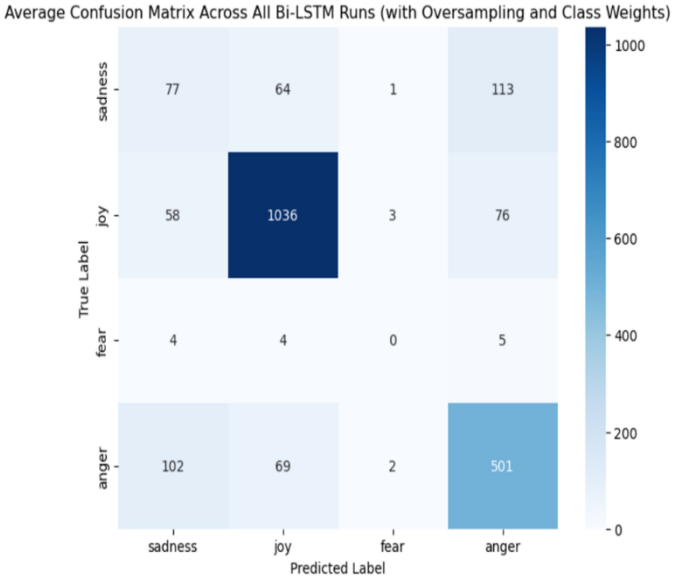
The evaluation of deep learning models without implementing any data balancing techniques shows relatively high testing accuracy, strong performance on the majority emotion classes, particularly "joy" and "anger." Among all architectures, LSTM-CNN and CNN yielded the highest mean test accuracies of 0.7960 and 0.7952, respectively, whereas macro F1-scores indicating inadequate performance on minority classes. The "fear" class shown zero F1-score across the models, and the "sadness" class also performed very poorly which highlights the effect of class imbalance, where the larger classes shows better accuracy but poor results for smaller classes. Conversely, when data balancing techniques (oversampling and class weighting) were applied, results show mean test accuracy for most models saw a slight decrease which indicates a tradeoff with overall predictive power but the Macro F1-scores significantly improved in all models. Based on the evaluation metrics, the best model without data balancing in terms of both accuracy and macro F1-score is the LSTM-CNN, achieving the highest accuracy of 79.60% and a macro F1-score of 0.4216. When data balancing is applied, the CNN model achieves the highest accuracy of 77.79%**,** while the Bi-LSTM model achieves the highest macro F1-score of 0.4978, indicating its superior performance in handling class imbalance and maintaining balanced precision and recall across all classes.

Table 4 presents a comparative analysis between existing models and the proposed model across various code-mixed datasets, including SemEval 2020 Task 9 (SentiMix), EmoMix-3L, and SemEval 2024 Task 10. Analysis highlights the various approaches and its performance in ED for bi-lingual CM data, specifically on CM or CS Hinglish text. Transformer-based models (e.g., XLMR, CM-RFT) on SemEval datasets generally show higher accuracy (e.g., 63-66%), some studies achieve comparable or even superior results (e.g., 71-83% accuracy with BERT, CNN, or CNN-BiLSTM) on custom or online datasets. On the SentiMix dataset, the proposed LSTM-CNN model on unbalanced data yields 79.60% accuracy and 0.4216 Macro F1-score. This accuracy is notably higher than the 66.03% accuracy of Transfer Learning-XLMR [14] 63.73% of the CM-RFT framework [18], and 66% reported by XLM-RoBERTa (XLM-R) [25] on the same dataset. This suggests that the CNN-LSTM hybrid architecture, even on unbalanced data, can effectively capture features relevant to emotion detection. Also, for balanced data CNN achieved highest performance of 77.79% accuracy and Bi-LSTM with highest Macro F1-score 0.4978.

Table 3. Performance of DL Models Based on Mean Test Accuracy and Mean Macro F1-score with and without data balancing

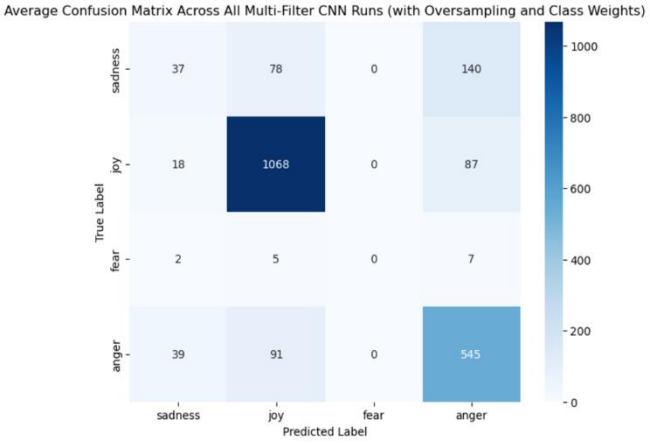
Model	Data Balance	Mean Test Accuracy (± 95% CI)	Mean Macro F1-score (± 95% CI)
Bi-LSTM	Unbalanced	76.34% ± 0.50%	0.4209 ± 0.0074
	Balanced	76.14% ± 0.70%	0.4978 ± 0.0226
CNN	Unbalanced	79.52% ± 0.43%	0.4178 ± 0.0026
	Balanced	77.79% ± 0.71%	0.4604 ± 0.0115
LSTM	Unbalanced	79.39% ± 0.24%	0.4177 ± 0.0015
	Balanced	51.33% ± 8.62%	0.1787 ± 0.0325
CNN-LSTM	Unbalanced	78.70% ± 0.53%	0.4137 ± 0.0025
	Balanced	74.79% ± 1.69%	0.4776 ± 0.0096
CNN-LSTM-CNN	Unbalanced	78.91% ± 0.53%	0.4148 ± 0.0029
	Balanced	76.25% ± 0.58%	0.4745 ± 0.0078
LSTM-CNN	Unbalanced	79.60% ± 0.28%	0.4216 ± 0.0040
	Balanced	75.57% ± 0.17%	0.4863 ± 0.0079

```
==================================================
Aggregated Results Across All Bi-LSTM Runs (with Oversampling and Class Weights)
==================================================
Mean Test Accuracy: 0.7614 ± 0.0070 (95% CI)
Standard Deviation of Test Accuracy: 0.0080
Mean Macro F1-score: 0.4978 ± 0.0226 (95% CI)
Standard Deviation of Macro F1-score: 0.0257

Per-class F1-scores (Mean across runs):
  sadness: 0.3084
  joy: 0.8821
  fear: 0.0705
  anger: 0.7303
```



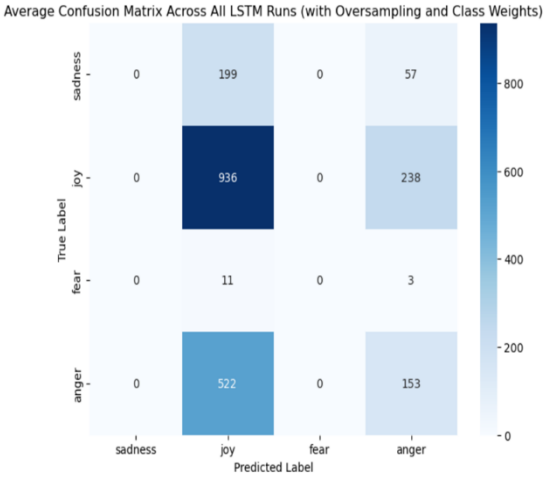
(a) Bi-LSTM

```
==================================================
Aggregated Results Across All Multi-Filter CNN Runs (with Oversampling and Class Weights)
==================================================
Mean Test Accuracy: 0.7779 ± 0.0071 (95% CI)
Standard Deviation of Test Accuracy: 0.0081
Mean Macro F1-score: 0.4604 ± 0.0115 (95% CI)
Standard Deviation of Macro F1-score: 0.0131

Per-class F1-scores (Mean across runs):
  sadness: 0.2092
  joy: 0.8835
  fear: 0.0000
  anger: 0.7489
```



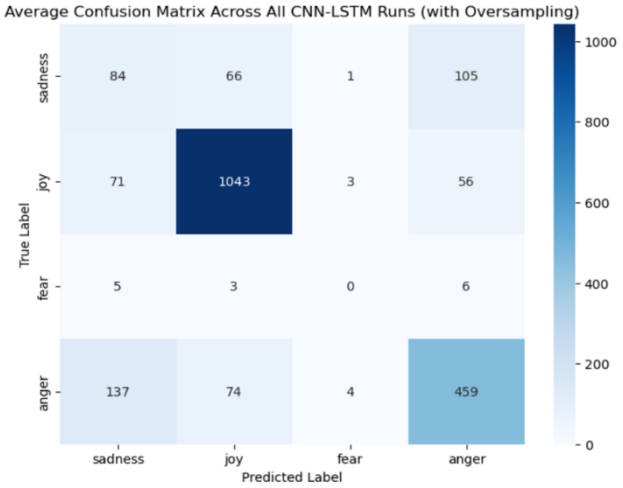
(b) CNN

```
==================================================
Aggregated Results Across All LSTM Runs (with Oversampling and Class Weights)
==================================================
Mean Test Accuracy: 0.5133 ± 0.0862 (95% CI)
Standard Deviation of Test Accuracy: 0.0984
Mean Macro F1-score: 0.1787 ± 0.0325 (95% CI)
Standard Deviation of Macro F1-score: 0.0371

Per-class F1-scores (Mean across runs):
  sadness: 0.0000
  joy: 0.5744
  fear: 0.0000
  anger: 0.1403
```



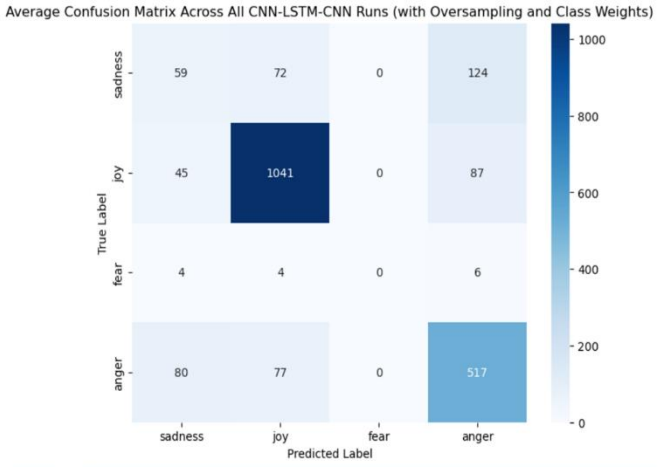
(c) LSTM

```
==================================================
Aggregated Results Across All CNN-LSTM Runs (with Oversampling)
==================================================
Mean Test Accuracy: 0.7479 ± 0.0169 (95% CI)
Standard Deviation of Test Accuracy: 0.0193
Mean Macro F1-score: 0.4776 ± 0.0096 (95% CI)
Standard Deviation of Macro F1-score: 0.0109

Per-class F1-scores (Mean across runs):
  sadness: 0.2904
  joy: 0.8836
  fear: 0.0340
  anger: 0.7023
```



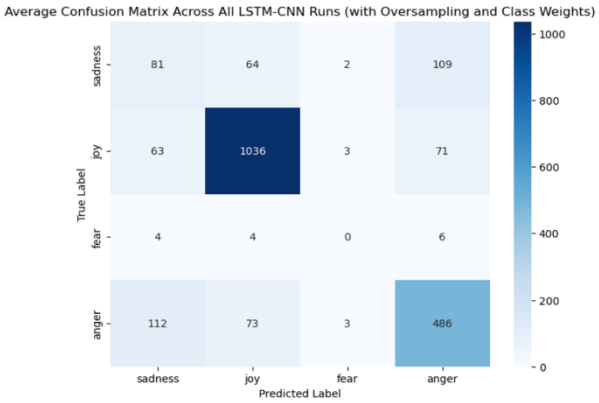
(d) CNN-LSTM

```
==================================================
Aggregated Results Across All CNN-LSTM-CNN Runs (with Oversampling and Class Weights)
==================================================
Mean Test Accuracy: 0.7625 ± 0.0058 (95% CI)
Standard Deviation of Test Accuracy: 0.0066
Mean Macro F1-score: 0.4745 ± 0.0078 (95% CI)
Standard Deviation of Macro F1-score: 0.0089

Per-class F1-scores (Mean across runs):
  sadness: 0.2659
  joy: 0.8789
  fear: 0.0211
  anger: 0.7323
```



(e) CNN-LSTM-CNN

```
==================================================
Aggregated Results Across All LSTM-CNN Runs (with Oversampling and Class Weights)
==================================================
Mean Test Accuracy: 0.7557 ± 0.0017 (95% CI)
Standard Deviation of Test Accuracy: 0.0020
Mean Macro F1-score: 0.4863 ± 0.0079 (95% CI)
Standard Deviation of Macro F1-score: 0.0090

Per-class F1-scores (Mean across runs):
  sadness: 0.3125
  joy: 0.8808
  fear: 0.0314
  anger: 0.7207
```



(f) LSTM-CNN

Figure 3. Aggregated results and Confusion matrix of performance of DL models with balanced data (a)Bi-LSTM, (b) CNN, (c) LSTM, (d) CNN-LSTM, (e) CNN-LSTM-CNN, (f) LSTM-CNN

Table 4. Comparative analysis of existing models and the proposed model on various code-mixed datasets such as SentiMix, EmoMix-3L and SemEval 2024 Task 10 dataset

References	Type of Language	Dataset	Performances (Accuracy in % and F1 Score)
[11]	Bi-Lingual (Code-Mixed)	SemEval-2024 as Task 10 Dialogue dataset	LSTM Model 37.8
[14]	Bi-Lingual (Code-Mixed)	SentiMix	Transfer Learning-XLMR: 66.03, F1score 0. 6447
[18]	Bi-Lingual (Code-Mixed)	SentiMix	CM-RFT framework 63.73 and F1 score 0.6053
[21]	Bi-Lingual (Code-Mixed)	EmoMix-3L	MuRIL 0.67 (F1 Score)
[23]	Bi-Lingual (Code-Mixed)	SemEval 2024 Task 10	Hing RoBERTa 47
[25]	Bi-Lingual (Code-Mixed)	SentiMix	XLM-RoBERTa (XLM-R) 66%, F1 score 0.72
[33]	Bi-Lingual (Code-Mixed)	Own Dataset	BERT 71.44
[36]	Bi-lingual (Code-Mixed)	Online (28667 tweets with 6 emotions)	CNN 75.25
[13]	Bi-lingual (Code-Mixed)	Own Dataset (12000 tweets with 3 emotions)	CNN-BiLSTM 83.21
Proposed Model	Bi-Lingual (Code-Mixed) Unbalanced Data	SentiMix	LSTM-CNN 79.60 and Macro F1-score 0.4216
Proposed Model	Bi-Lingual (Code-Mixed) Balanced Data	SentiMix	CNN 77.79**,** Bi-LSTM Macro F1-score 0.4978

5. Conclusion

Proposed study gives contribution in the area of ED by analyzing various DL models like Bi-LSTM, CNN, LSTM, and hybrid models such as CNN-LSTM, LSTM-CNN and CNN-LSTM-CNN on CM and CS social media text. The results show the hybrid models especially LSTM-CNN model, achieve higher performance by taking advantages of both convolutional and recurrent architectures for ED. Our findings focus on benefit of used of combining CNN and LSTM architectures for ED from CM text. This study provides valuable insights into various DL approaches used for ED in social media text. Future research can focus on improving model's efficiency using various DL techniques or attention mechanisms, knowledge graphs etc. Also, availability of annotated CM datasets in different languages is still limited.

References

[1] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu, V. Varadarajan, AI Based Emotion Detection for Textual Big Data: Techniques and Contribution. Big Data and Cognitive Computing, 5(3), (2021) 43. https://doi.org/10.3390/bdcc5030043

[2] A. Al Maruf, F. Khanam, M.M. Haque, Z.M. Jiyad, M.F. Mridha, Z. Aung, Challenges and Opportunities of Text-Based Emotion Detection: A Survey. IEEE Access, IEEE, 12, (2024)18416–18450. https://doi.org/10.1109/ACCESS.2024.3356357

[3] F.A. Acheampong, C. Wenyu, H. Nunoo-Mensah, Text-based emotion detection: Advances, challenges, and opportunities. Engineering Reports, 2(7), (2020) e12189. https://doi.org/10.1002/eng2.12189

[4] P. Nandwani, R. Verma, A review on sentiment analysis and emotion detection from text. Social Network Analysis Mining, 11(1), (2021) 81. https://doi.org/10.1007/s13278-021-00776-6

[5] P. Ekman, Basic Emotions. Handbook of Cognition and Emotion, (1999) 45–60. https://doi.org/10.1002/0470013494.ch3

[6] R. Jan, A.A. Khan, Emotion Mining Using Semantic Similarity. International Journal of Synthetic Emotions, 9(2), (2018) 1–22.

[7] S.A. Kumar, A. Geetha, Emotion Detection from Text using Natural Language Processing and Neural Networks. International Journal of Intelligent Systems and Applications in Engineering, 12(14), (2024) 609–615. https://ijisae.org/index.php/IJISAE/article/view/4707

[8] J. Guo, Deep learning approach to text analysis

for human emotion detection from big data. Journal of Intelligent Systems, 31(1), (2022)113–126. https://doi.org/10.1515/jisys-2022-0001

[9] D. Yohanes, J.S. Putra, K. Filbert, K.M. Suryaningrum, H.A. Saputri, Emotion Detection in Textual Data using Deep Learning. Procedia Computer Science, 227, (2023) 464–473, 2023 https://doi.org/10.1016/j.procs.2023.10.547

[10] S. Mubeen, N. Kulkarni, M.R. Tanpoco, R.D. Kumar, L.M. Naidu, T. Dhope, Linguistic Based Emotion Detection from Live Social Media Data Classification Using Metaheuristic Deep Learning Techniques. International Journal of Communication Networks and Information Security (IJCNIS), 14(3), (2022) 176–186. https://doi.org/10.17762/ijcnis.v14i3.5604

[11] V. Ravindran, A. Jetti, R. Sivanaiah, A. Deborah, M. Thankanadar, R.S. Milton, (2024) TECHSSN at SemEval-2024 Task 10: LSTM-based Approach for Emotion Detection in Multilingual Code-Mixed Conversations. Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), pp. 763–769. https://doi.org/10.18653/v1/2024.semeval-1.109

[12] K. Yadav, A. Lamba, D. Gupta, A. Gupta, P. Karmakar, S. Saini, (2020) Bi-LSTM and Ensemble based Bilingual Sentiment Analysis for a Code-mixed Hindi-English Social Media Text. 2020 IEEE 17th India Council International Conference (INDICON), IEEE, New Delhi, India. https://doi.org/10.1109/INDICON49873.2020.9342241

[13] T.T. Sasidhar, B. Premjith, K.P. Soman, Emotion Detection in Hinglish (Hindi+English) Code-Mixed Social Media Text. Procedia Computer Science, 171, (2020) 1346–1352. https://doi.org/10.1016/j.procs.2020.04.144

[14] S. Ghosh, A. Priyankar, A. Ekbal, P. Bhattacharyya, Multitasking of sentiment detection and emotion recognition in code-mixed Hinglish data. Knowledge Based System, 260, (2023)110182. https://doi.org/10.1016/j.knosys.2022.110182

[15] U. Barman, A. Das, J. Wagner, J. Foster, Code Mixing: A Challenge for Language Identification in the Language of Social Media. 1st Workshop on Computational Approaches to Code Switching, Switching 2014 at the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP (2014) 13–23. https://doi.org/10.3115/v1/W14-3902

[16] K.K. Sampath, M. Supriya, Transformer Based Sentiment Analysis on Code Mixed Data. Procedia Computer Science, 233, (2024) 682–691. https://doi.org/10.1016/j.procs.2024.03.257

[17] P. Udawatta, I. Udayangana, C. Gamage, R. Shekhar, S. Ranathunga, Use of prompt-based learning for code-mixed and code-switched text classification, World Wide Web, 27(5) ,(2024) 63. https://doi.org/10.1007/s11280-024-01302-2

[18] G.V. Singh, S. Ghosh, M. Firdaus, A. Ekbal, P. Bhattacharyya, Predicting multi-label emojis, emotions, and sentiments in code-mixed texts using an emojifying sentiments framework. Scientific Reports, 14(1), (2024) 12204. https://doi.org/10.1038/s41598-024-58944-5

[19] A. Thiab, L. Alawneh, M.AL-Smadi, Contextual emotion detection using ensemble deep learning. Computer Speech Language, 86, (2024) 101604. https://doi.org/10.1016/j.csl.2023.101604

[20] R. Agarwal, N. Abbas, (2024) Emotion Detection in Hindi Language Using GPT and BERT. In International Conference on Innovative Techniques and Applications of Artificial Intelligence, 105–118. https://doi.org/10.1007/978-3-031-77918-3_8

[21] N. Raihan, D. Goswami, A. Mahmud, A. Anastasopoulos, M. Zampieri, (2024). EmoMix-3L: A Code-Mixed Dataset for Bangla-English-Hindi Emotion Detection. arXiv. https://doi.org/10.48550/arXiv.2405.06922

[22] H. Takahashi, (2024) Hidetsune at SemEval-2024 Task 10: An English Based Approach to Emotion Recognition in Hindi-English code-mixed Conversations Using Machine Learning and Machine Translation. In Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024) 374-378. https://doi.org/10.18653/v1/2024.semeval-1.58

[23] V.O. Yenumulapalli, P. Premnath, P. Mohankumar, R. Sivanaiah, A. Deborah, (2024) TECHSSN1 at SemEval-2024 Task 10: Emotion Classification in Hindi-English Code-Mixed Dialogue using Transformer-based Models. In Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), 833-838. https://doi.org/10.18653/v1/2024.semeval-1.119

[24] L. S. Meetei, S. M. Singh, A. Singh, R. Das, T. D. Singh, S. Bandyopadhyay, Hindi to English multimodal machine translation on news dataset in low resource setting. Procedia Computer Science, 218, (2023) 2102-2109. https://doi.org/10.1016/j.procs.2023.01.186

[25] M. Imam, A. Patnaik, S. Jena, M. Digdarshini, S. K. Sharma, D. Muduli, Integrated Approach for Sentiment Detection and Emotion Recognition in Code-Mixed Hinglish Data. International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), IEEE, India. https://doi.org/10.1109/SCOPES64467.2024.10990930

[26] S. Mishra, S. Suryavardan, M. Chakraborty, P. Patwa, A. Rani, A. Chadha, A. Reganti, A. Das, A. Sheth, M. Chinnakotla, A. Ekbal, S. Kumar, (2023) Overview of memotion 3: Sentiment and

emotion analysis of codemixed hinglish memes. arXiv. https://doi.org/10.48550/arXiv.2309.06517

[27] J. Herzig, M. Shmueli-Scheuer, D. Konopnicki, (2017) Emotion detection from text via ensemble classification using word embeddings. ICTIR 2017 - Proceedings of the 2017 ACM SIGIR International Conference on the Theory of Information Retrieval, 269–272. https://doi.org/10.1145/3121050.3121093

[28] M. Polignano, M. De Gemmis, P. Basile, G. Semeraro, (2019) A Comparison of Word-Embeddings in Emotion Detection from Text using BiLSTM, CNN and Self-Attention.In Adjunct publication of the 27th conference on User Modeling, Adaptation, and Personalization, 63–68. https://doi.org/10.1145/3314183.3324983

[29] K. Shrivastava, S. Kumar, D.K. Jain, An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network. Multimedia Tools Applications, 78, (2019) 29607–29639. https://doi.org/10.1007/s11042-019-07813-9

[30] A. Perera, A. Caldera, Sentiment Analysis of Code-Mixed Text: A Comprehensive Review.Journal of Universal Computer Science, 30(2), (2024) 242–261. https://doi.org/10.3897/jucs.98708

[31] P. Patwa, G. Aguilar, S. Kar, S. Pandey, S. Pykl, B. Gambäck, T. Chakraborty, T. Solorio, A. Das, (2020) SemEval-2020 Task 9: Overview of Sentiment Analysis of Code-Mixed Tweets. 14th International Workshops on Semantic Evaluation, SemEval 2020 - co-located 28th International Conference on Computational Linguistics, 774–790. https://doi.org/10.18653/v1/2020.semeval-1.100

[32] D. Vijay, A. Bohra, V. Singh, S.S. Akhtar, M. Shrivastava, (2018) Corpus creation and emotion prediction for hindi-english code-mixed social media text. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Student Research Workshop, 128–135. https://doi.org/10.18653/v1/N18-4018

[33] A. Wadhawan, A. Aggarwal, (2021) Towards Emotion Recognition in Hindi-English Code-Mixed Data: A Transformer Based Approach, arXiv.

[34] B.G. Patra, D. Das, A. Das, (2018) Sentiment Analysis of Code-Mixed Indian Languages: An Overview of SAIL_Code-Mixed Shared Task @ICON-2017. ArXiv.

[35] P. Mishra, P. Danda, P. Dhakras, (2018) Code-Mixed Sentiment Analysis Using Machine Learning and Neural Network Approaches. arXiv.

[36] S. Das,T. Singh,(2023) Sentiment Recognition of Hinglish Code Mixed Data using Deep Learning Models based Approach. Proceedings of the 13th International Conference on Cloud Computing, Data Science and Engineering (Confluence), IEEE, Noida, India. 265–269. https://doi.org/10.1109/Confluence56041.2023.10048879

[37] A. Patil, V. Patwardhan, A. Phaltankar, G. Takawane, and R. Joshi, (2023) Comparative Study of Pre-Trained BERT Models for Code-Mixed Hindi-English Data. IEEE 8th International Conference for Convergence in Technology (I2CT), IEEE, Lonavla, India. https://doi.org/10.1109/I2CT57861.2023.10126273

[38] X. Zhu, Y. Lou, H. Deng, D. Ji, Leveraging bilingual-view parallel translation for code-switched emotion detection with adversarial dual-channel encoder. Knowledge Based Systems, 235, (2022) 107436. https://doi.org/10.1016/j.knosys.2021.107436

[39] H. Rathnayake, J. Sumanapala, R. Rukshani, S. Ranathunga, AdapterFusion-based multi-task learning for code-mixed and code-switched text classification. Engineering Applications of Artificial Intelligence, 127, (2024) 107239. https://doi.org/10.1016/j.engappai.2023.107239

[40] A.R. Abas, I. Elhenawy, M. Zidan, M. Othman, BERT-CNN: A Deep Learning Model for Detecting Emotions from Text. Computers, Materials and Continua, 71(2), (2021) 2943–2961. https://doi.org/10.32604/cmc.2022.021671

[41] F.M.P. Del Arco, A. Curry, A.C. Curry, D. Hovy, (2024) Emotion Analysis in NLP: Trends, Gaps and Roadmap for Future Directions. International Conference on Language Resources and Evaluation, arXiv. https://doi.org/10.48550/arXiv.2403.01222

[42] A.N.B. Emran, A. Ganguly, S.S.C. Puspo, N. Raihan, D. Goswami, (2024) MasonTigers at SemEval-2024 Task 10: Emotion Discovery and Flip Reasoning in Conversation with Ensemble of Transformers and Prompting. arXiv. https://doi.org/10.48550/arXiv.2407.00581

[43] S. Kumar, M.S. Akhtar, E. Cambria, T. Chakraborty, (2024) SemEval 2024 -- Task 10: Emotion Discovery and Reasoning its Flip in Conversation (EDiReF), arXiv. https://doi.org/10.48550/arXiv.2402.18944

Acknowledgement

We would like to acknowledge the online platforms such as Kaggle, GitHub for availability of online dataset and preparation of dictionary used in this study. Both the authors read and agreed the final version of the manuscript.

Authors Contribution Statement

Vinayak Malavade: Conceptualization, data collection, data analysis, writing original manuscript. Virat Giri: Guidance for implementation, paper writing and submission, reviews and editing the text.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Has this article screened for similarity?

Yes

About the License