International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 402-411 March 2026-Sept 2026 www.csjournals.com

Sentiment and Emotions Detection in Social Media Text by using ML and Transformer-Based Techniques

Swati, Dr. Naveen and Dr. Anupam Bhatia

MCA Student[1], Asst. Professor[2], Associate Prof.[3]

Deptt. of Comp. Sc. and Application, Ch. Ranbir Singh University, Jind, Haryana, India

ABSTRACT

The detection of sentiment and emotions in the Hindi-English mixed social media texts is not easy because of code-mixing, transliteration differences, informal language usage, and class imbalance. In this research, we compare Classical ML models and Transformer-based models for sentiment classification (positive, negative and neutral) and emotion detection (anger, disgust, fear, joy, sadness, surprise etc.) on the SentiMix-EmoSen dataset comprising 20,000 instances. The comparison is made between Classical models such as NB , LR, Linear SVM, and Dual-View SVM against multilingual Transformer models like Distil-mBERT, mBERT, XLMRoBERTa, Twitter-XLM-RoBERTa, and a custom XLM-R Multitask model trained via class-weighted learning. According to the experimental results, the Twitter-XLM-RoBERTa provides the best sentiment classification accuracy with 73.40% and the weighted F1-score of 73.34%, performing better than the best classical model (Linear SVM) by 4.72 percentage points. The XLM-R Multitask model provides the best emotion detection performance with the accuracy of 64.97% and the weighted F1-score of 64.26%. However, the Distil-mBERT performs competitively due to its smaller architecture. The analysis also finds out that the Fear and Surprise classes are hard to detect because of significant class imbalance.

Keywords: Sentiment Analysis, Emotion Recognition, Hinglish, Code-Mixed Text, ML, Transformer Models, Multitask Learning, XLM-RoBERTa, SentiMix-EmoSen, Natural Language Processing.

Abbrevations

- Term Frequency–Inverse Document Frequency - TF-IDF
- Machine Learning - ML
- Natural Language Processing - NLP
- Logistic Regression - LR
- Naïve Bayes -NB
- Support Vector Machine - SVM
- Decision Tree - DT
- Random Forest - RF
- Multilingual BERT - mBERT

INTRODUCTION

The fast-paced development of social media sites has changed the methods used for communication, exchanging ideas and expressing feelings. Sites like Twitter (X), Facebook, Instagram, and Reddit produce massive amounts of textual data produced by users that express opinions about politics, healthcare, education, businesses, entertainment, and social matters. Such a flow of information has generated new ways of receiving useful insights about online discussions, thus making opinion mining a relevant topic in NLP [2], [3].

The two most popular tasks in opinion mining are the Sentiment Analysis and Emotion Recognition due to their capability of automated interpretation of opinions and emotions. The first one analyses the polarity of a text by determining whether the opinion is expressed positively, negatively or neutrally, while Emotion Recognition uncover specific emotions like anger, disgust, fear, joy, sadness, surprise and other [4], [5]. These techniques have been successfully implemented for various purposes such as : customer satisfaction analysis, recommendation systems, health monitoring, finance prediction, political opinion mining and brand reputation management.

The Sentiment and Emotion Analysis accuracy mainly relies on the Natural NLP technology. It is one of the branches of artificial intelligence used for processing the languages of humans. Preprocessing methods including text cleaning, tokenization, normalization, stop-word elimination, lemmatization, and feature extraction allow converting raw textual data into some meaningful representation [2], [5]. Nevertheless, social media text analysis is a difficult task due to the use of informal language, abbreviations, emoticons, hashtags, misspellings, transcribed words, and multilingualism and code-mixing in one expression [4], [6]–[10].

ML has emerged as one of the most widely utilized methods to perform automated text classification to overcome those obstacles. Classical ML algorithms such as NB, LR, DT, RF and SVM are often used for analyzing sentiments because of its efficiency and potential to process multidimensional textual attributes obtained by such methods as TF-IDF [2], [3], [9], [13] Despite their good performance, those models often do not take into account contextual information and long-term relationships inherent to multilingual and code-mixed social media text [7]–[14].

Recently, there have been many breakthroughs in Deep Learning in general and language modeling in particular, with Transformer-based models leading to better results in sentiment/emotion analysis. Models such as BERT, DistilBERT, mBERT, and XLM-RoBERTa employ self-attention mechanisms to generate contextual word representations by considering relationships among all words in a sentence. These contextual embeddings enable better understanding of multilingual semantics and long-range contextual dependencies, making Transformer models particularly effective for analysing Hinglish code mixed social media text [5], [10]–[12], [14]. Despite their superior performance, challenges such as computational complexity, domain variability, and severe class imbalance continue to limit their effectiveness, especially in low-resource multilingual environments [7], [8], [10]–[12], [14]. Although several studies have explored either sentiment analysis or emotion recognition for multilingual and code-mixed text, comprehensive evaluation of classical ML and Transformer-based multilingual models under a common experimental framework remains limited [5], [11], [12]. Furthermore, relatively few studies jointly investigate both sentiment classification and fine-grained emotion recognition using the same Hinglish dataset while analysing the influence of the class imbalance on the model effectiveness. Motivated by these challenges, this research provides a detailed comparative evaluation of classical ML and Transformer based multilingual models for sentiment categorization and emotion recognition using the SentiMix-EmoSen dataset comprising approximately 20,000 Hinglish code-mixed social media posts. Classical algorithms including Naïve Bayes, Logistic Regression, Linear SVM, and Dual-view SVM are evaluated alongside multilingual Transformer models such as Distil mBERT, mBERT, XLM-RoBERTa, and Twitter-XLM-RoBERTa, together with a dedicated XLM-R Multitask model employing class-weighted learning.

The key findings of this study are summarized as:

- A comparative study of classical ML and Transformer based multilingual models for sentiment evaluation and emotion identification of Hinglish mixed linguistic social media text.

- Development and evaluation of a dedicated XLM-R Multitask model for jointly learning sentiment representations. and emotion

- In-depth experimental evaluation with SentiMix-EmoSen dataset with Accuracy and Weighted F1-score as main performance indicators.

- Investigation of the impact of class imbalance on emotion recognition together with discussion of potential strategies for improving multilingual code mixed NLP systems.

Experimental evaluation demonstrates that Twitter XLM-RoBERTa achieves the best sentiment classification performance, while the proposed XLM-R Multitask model provides the highest emotion recognition performance. The comparative analysis confirms the superiority of Transformer-based multilingual models over conventional ML approaches and highlights the importance of contextual representation learning, multitask learning, and class-balancing strategies for improving sentiment and emotion review of a Hinglish code mixed online social media content.

Related Work

A. Early Sentiment Evaluation on Mixed Language Text

Early study on Hinglish sentiment evaluation established SVM with TF-IDF features as one of the strongest classical baselines for code-mixed text classification [2], [9]. The development of annotated code-mixed corpora further accelerated multilingual sentiment analysis by providing standardized benchmarks for model evaluation [4]. Subsequent studies on multilingual and Tamil code-mixed sentiment analysis demonstrated that traditional ML models combined with lexical and character-level features remain effective for low-resource languages, motivating the use of TF-IDF-based Linear SVM and the proposed Dual-view SVM architecture in this work [4], [10], [11].

B. Transformer-Based Sentiment Analysis

Today's progress in transformer-based linguistic models has drastically enhanced multilingual sentiment analysis by learning contextual representations through self-attention mechanisms. Multilingual models such as mBERT, DistilBERT, XLM-RoBERTa, and domain-adapted Transformer architectures consistently outperform traditional ML approaches on multilingual and code-mixed datasets [5], [10]–[12]. Hashmi et al. demonstrated that multilingual Transformers substantially improve sentiment prediction for code-mixed tweets, while Phadte and Dhore validated the effectiveness of multilingual Transformer models for social media sentiment analysis [10], [12]. Similarly, Zhu reported that Transformer-based models achieve higher classification accuracy for Twitter sentiment analysis by effectively capturing contextual information from social media text [6].

C. Emotion Identification in Resource-Poor Languages

Emotion identification in multilingual and poor resource languages remains more challenging than sentiment classification because of class imbalance, transliteration ambiguity, and contextual dependency. Ullah et al. presented a CNN-LSTM infrastructure for Urdu emotion detection and identified class imbalance as a major factor affecting model performance [8]. Nagra et al. demonstrated that deep learning models effectively capture transliterated text representations for Roman Urdu sentiment analysis [7]. Likewise, Chakravarthi et al. highlighted annotation challenges arising from sarcasm and contextual ambiguity in

code-mixed datasets, while Gowda et al. improved Tamil code mixed sentiment analysis using meta-model ensembling techniques [4], [11].

D. Multitask Learning for Affective Computing

It has been observed that multitask learning is capable of enhancing the functionality of analysis of sentiment and detection of emotions by learning shared semantic representations at the same time. Ghosh and coworkers suggested a multi-task learning approach for sentiment and emotion analysis in Hinglish social media messages, which is proven to be effective for both tasks [5]. Phadte and Dhore further validated multilingual approaches for code-mixed social media sentiment analysis [10]. More recently, Prova et al. employed GPT and a weighted ensemble deep neural network framework for multi-lingual sentiment estimation, achieving improved performance across multiple languages [14]. Furthermore, Aggarwal and Sharma & Singh confirmed that well-optimized SVM models with TF-IDF features remain strong and computationally efficient baselines when labelled training data are limited [2], [13].

E. Research Gap

The aforementioned gaps in research are recognized based on the existing literature examination:

- Almost existing studies focus on either sentiment analysis or emotion recognition, with limited work addressing both tasks under a unified framework.
- Comparative evaluation of classical ML, general-purpose multilingual Transformers, and domain specific Transformer models for Hinglish code-mixed text remains limited.
- Severe class imbalance, particularly for minority emotions such as Fear and Surprise, continues to reduce the performance of recognition models.
- Limited studies existing emotion investigate the effectiveness of lightweight multilingual Transformers compared with larger Transformer architectures for code-mixed sentiment and emotion analysis.
- A comprehensive evaluation using a common benchmark dataset with standardized performance metrics is still lacking.

Table 1: Comparison Studies of Existing Research

Authors & Year	Methodology	Key Participation	Shortcomings
Sharma & Singh (2023)	Classical ML	Discussed the use of ML algorithms in everyday text classification tasks	Does not specifically address code-mixed social media text
Go et al. (2009)	Naïve Bayes, Maximum Entropy, SVM	Introduced distant supervision using emoticons for automatic sentiment labeling	Noisy labels due to emoticon ambiguity
Chakravarthi et al. (2020)	Corpus Annotation	Developed a benchmark Tamil-English code-mixed sentiment dataset	Limited to Tamil-English language pair
Ghosh et al. (2023)	BiLSTM Multi-task Learning	Jointly learned sentiment and emotion representations for Hinglish text	Requires large annotated datasets and longer training
Zhu (2023)	Transformer-based NLP	Demonstrated the effectiveness of contextual language models for Twitter sentiment analysis	Limited to political Twitter data
Nagra et al. (2022)	FRCNN (CNN-RNN Hybrid)	Improved sentiment classification for transliterated Roman Urdu text	High computational cost
Ullah et al. (2022)	CNN-LSTM	Enhanced sentiment and emotion detection for low-resource languages	Performance affected by class imbalance and limited training data
Singh	NLP and ML	Established classical Hinglish sentiment analysis using ML techniques	Limited dataset availability
Phadte & Dhore (2025)	Multilingual NLP	Extended sentiment analysis to multilingual Indian code-mixed languages	Transliteration and code-switching challenges
Gowda et al. (2025)	Meta-model	Improved sentiment prediction through	Increased model complexity

	Ensemble	ensemble learning	
Hashmi et al. (2024)	Multilingual Transformer Models	Improved contextual understanding and prediction for code-mixed tweets	Computationally expensive
Prova et al. (2026)	GPT + Weighted Ensemble	Improved multilingual sentiment prediction using GPT-based ensemble learning	Resource-intensive training and inference
Bandarupalli (2025)	Transformer-based NLP	Demonstrated superior multilingual sentiment classification using Transformer architectures	High computational requirements
Singhal & Bishnoi (2023)	TF-IDF + Support Vector Machine	Established Hinglish sentiment classification using classical ML techniques	Limited dataset diversity and scalability

RESEARCH METHODOLOGY

This proposed framework aims at performing classification of sentiment and emotion identification for Hindi English (Hinglish) mixed linguistic digital media content through an interdisciplinary examination of classical ML and Transformer based multilingual modeling techniques. The overall methodology comprises dataset gathering, preprocessing of text, characteristic representation, modeling and training, multitask learning and performance assessment, as illustrated in Fig. 1..

A. Data Preparation

Experiments are performed on the SentiMix-EmoSen dataset consisting of about 20k Hinglish social media posts annotated with three sentiment classes (positive, negative, and neutral) and seven emotion categories (anger, disgust, fear, joy, sadness, surprise, and others). This dataset was chosen since it reflects natural multilingual conversations on social media with code-switching, transliteration, spelling errors, and other informality features and, thus, serves as a good evaluation benchmark for multilingual sentiment analysis methods [2], [3].

Table 2: Dataset Attributes

Attribute	Description
id	Unique identifier
tweet	Social media text
sentiment	Sentiment label
emotion	Emotion label
meta	Additional metadata

Table 3: Overall Distribution

Full Dataset Loaded: Total Records: 20000	
Sentiment	**Emotion**
Neutral 7492 Positive 6616 Negative 5892	Others 8114 Joy 5874 Anger 3382 Sadness 1285 Disgust 1208 Fear 73 Surprise 64

International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 402-411 March 2026-Sept 2026 www.csjournals.com

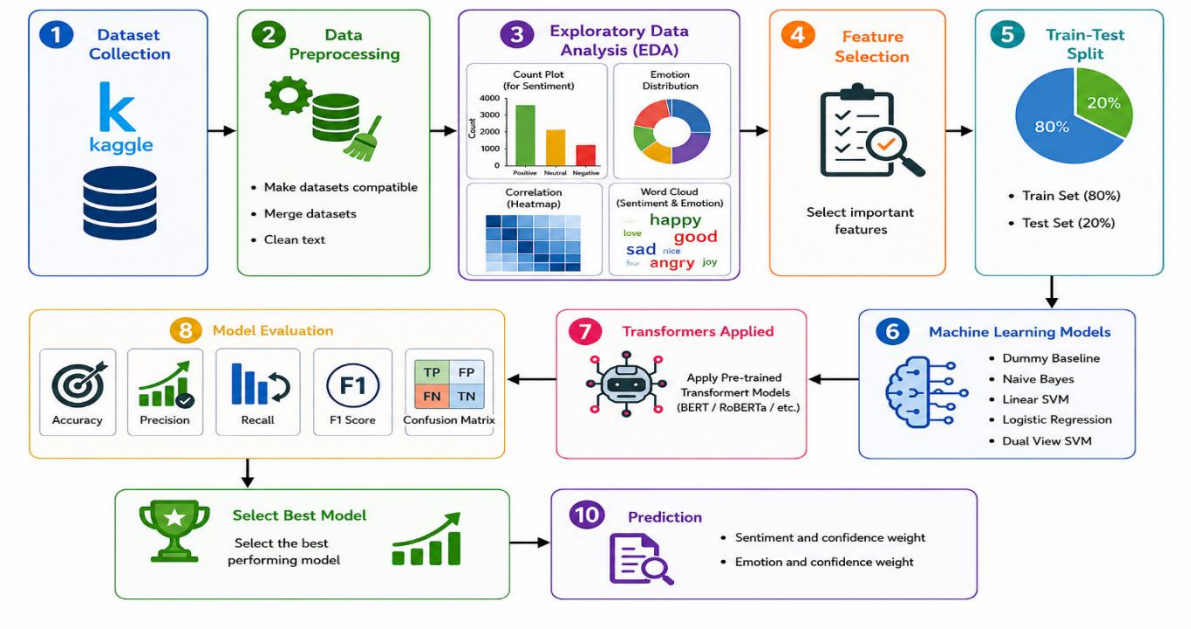


Fig 1: Proposed Architecture

B. Data Preprocessing

The nature of social media text is noisy as it includes URLs, hashtags, user mentions, emojis, character repetitions, punctuation marks, and misspellings. In order to preprocess the collected data, text normalization, converting into lower case, removing of URLs, user mentions, punctuation, and stopwords, and then tokenization were performed. Such preprocessing steps decrease textual noise and preserve semantic meaning [3], [9].

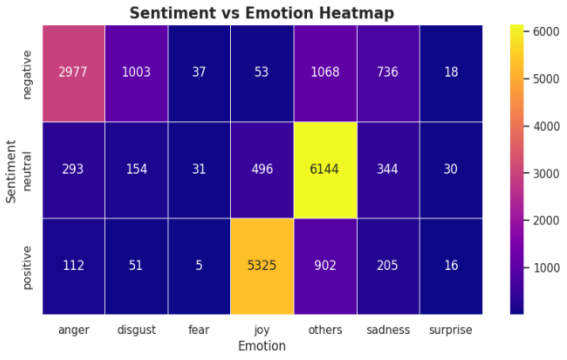
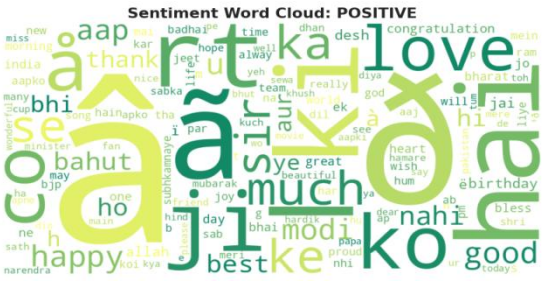


Fig 2: Sentiment vs Emotion Heatmap



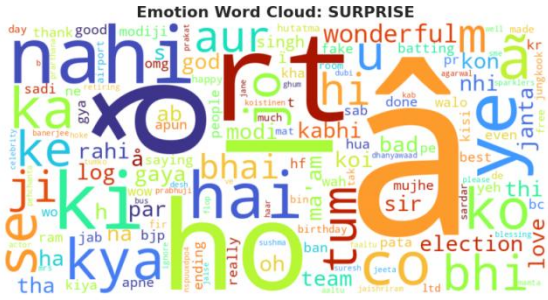


Fig 3: Word Cloud for Sentiment and Emotion

C. Feature Representation and Classification

Two types of paradigms were used to perform sentiment and emotion classification: classical ML paradigm and Transformer-based multilingual language models.

1) Classical ML Models

Five classical ML algorithms were considered to determine the appropriate baseline performance of sentiment and emotion classification. The chosen algorithms are known for being computationally effective, interpretable and applied widely for solving tasks related to text classification. Textual data were represented using TF-IDF, which transforms textual knowledge into quantitative vectors of features, giving higher priority to useful and informative words and less generally occurring phrases. [7], [8].

a) Dummy Baseline

The Dummy Baseline is a simple reference classifier that predicts labels using predefined strategies, such as selecting the most frequent class or generating random predictions,

without learning any relationship from the input features. Although it does not perform actual learning, it provides a minimum performance benchmark against which all trained models can be compared. In this research, the Dummy Baseline was used to verify that the proposed ML and Transformer models learn meaningful linguistic patterns rather than producing results comparable to random or majority-class predictions.

b) Naïve Bayes (NB)

NB is a supervised deterministic classification approach built around Bayes' theorem that specifies that attributes are unconditionally self-sufficient based on the target class Although this is a simplified hypothesis, NB works very well when dealing with high dimensional text classification problems because word occurrences can be efficiently modelled using probability distributions. In this research, NB was selected as a lightweight baseline to evaluate the effectiveness of probabilistic learning on TF-IDF feature vectors extracted from Hinglish code-mixed social media text.

c) Logistic Regression (LR)

LR is a supervised linear classifier technique, which uses the logistic (sigmoid) function to estimate the probable outcome of each category. Unlike probabilistic generative models, LR directly learns the relationship between textual features and class labels by optimizing a determination threshold that categorizes particular cases. It was included in this research because of its robustness, computational efficiency, and proven effectiveness for sparse text representations generated through TF-IDF feature extraction.

d) Linear Support Vector Machine (Linear SVM)

Linear SVM is a discriminative learning technique that learns the optimum hyperplane that distinguishes various categories and maximizes the separation among each of them. Because TF-IDF representations are high dimensional and sparse, Linear SVM has consistently demonstrated superior performance for text classification compared with many traditional ML algorithms. Therefore, Linear SVM was adopted as the primary classical baseline for both sentiment classification and emotion recognition in this study.

e) Dual-view Support Vector Machine (Dual-view SVM)

Dual-view SVM is an extension of the traditional Linear SVM that learns from two distinct feature representations. In contrast to the traditional Linear SVM, where only one textual representation is learned, the Dual-view SVM learns from two views of the input text which enables the capturing of both lexical and context-based features of Hinglish code-mixed language. Such an approach helps better recognize variations in spelling, transliterations, and colloquial forms used in social media texts.

2) Transformer-Based Models

To overcome the limitations of feature-based ML algorithms, four previously trained multilingual Transformer models were investigated for sentiment evaluation and emotion detectionThe Transformer model's main feature is attention to itself, which allows the system to learn the relationship between words, considering the entire input sequence at the same time. Unlike traditional ML methods that rely on handcrafted features, Transformer models automatically learn semantic and contextual representations, making them highly effective for multilingual and code-mixed NLP tasks [15]–[18].

a) Distil-mBERT

Distil-mBERT is a light-weight version of mBERT created by knowledge distillation, in which a more compact typical pupil model learns how to mimic the behavior of a higher-level teacher model. It contains significantly fewer parameters while retaining most of the language understanding capability of mBERT, resulting in faster training and lower computational requirements. We chose to investigate if a compact multilingual transformer might yield competitive results with a cut in computational cost.

b) Multilingual BERT (mBERT)

mBERT is a pre-trained Transformer model trained on multilingual Wikipedia corpora using the Masked Language Modelling (MLM) objective. It generates contextual word embeddings by taking into account the left and right circumstances surrounding each token, enabling effective cross-lingual knowledge transfer. Since Hinglish combines Hindi and English within the same sentence, mBERT was employed because of its ability to understand multilingual vocabulary and contextual semantics without requiring language specific feature engineering.

c) XLM-RoBERTa (XLM-R)

XLM-R is an enhanced multilingual Transformer trained on a large-scale multilingual corpus. Compared with mBERT, it provides stronger contextual representations and improved cross-lingual generalization. It was included because it has consistently demonstrated state-of-the-art results on multi-lingual sentiment evaluation and code-mixed NLP tasks, making it an effective benchmark for evaluating contextual representation learning [15], [16].

d) Twitter-XLM-RoBERTa

version of XLM-RoBERTa further fine-tuned on multilingual Twitter data. Unlike general-purpose language models, it learns the linguistic characteristics of social media text, including informal vocabulary, hashtags, emojis, abbreviations, spelling variations, and code-mixed expressions. Since the SentiMix-EmoSen dataset consists of Hinglish social media posts, this model was selected to

International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 402-411 March 2026-Sept 2026 www.csjournals.com

investigate whether domain-specific pre-training improves sentiment and emotion classification performance.

D. Multitask Learning Framework

Since sentiment and emotion are semantically related affective tasks, a dedicated XLM-R Multitask model was implemented to learn both tasks simultaneously. The model employs a shared XLM-R encoder with separate task-specific classification heads for sentiment and emotion prediction. Shared representation helps the knowledge transfer between tasks and helps generalize the features better while avoiding overfitting. To minimize the effect of severe class imbalance among emotion categories, class-weighted loss was incorporated during training, assigning higher importance to minority emotion classes [2].

E. Performance Evaluation Metrics

To comprehensively evaluate the performance of all ML and Transformer models, six standard evaluation metrics were employed.

1) Accuracy

Accuracy measures overall correctness.

Formula:

$$\textbf{Accuracy=\frac{TP+TN}{TP+TN+FP+FN}}$$

Where: TP = True Positive , FP = False Positive

TN = True Negative , FN = False Negative

2) Precision

Precision measures true positive predicted positive instances.

Formula: $$\textbf{Precision=\frac{TP}{TP+FP}}$$

3) Recall

Recall, also known as Sensitivity, measures correctly identified positive instances,

Formula: $$\textbf{Recall=\frac{TP}{TP+FN}}$$

4) Weighted F1-score

Harmonic mean of Precision and Recall.

Formula: $$\textbf{F1=\frac{2\times Precision\times Recall}{Precision+Recall}}$$

5) Confusion Matrix

Compares the actual instances with the predicted instances.

Results and Discussion

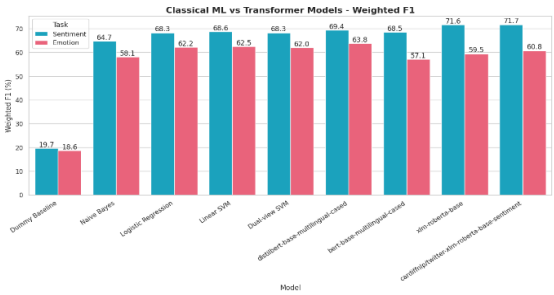


Fig 4: Classical ML vs Transformer Models: Weighted F1

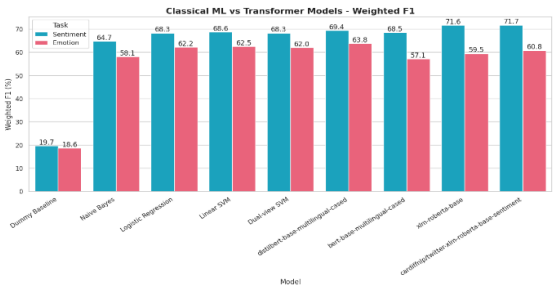


Fig 5: Classical ML vs Transformer Models: Accuracy

Linear SVM showed the best results among the classical ML algorithms with 68.97% of sentiment accuracy (68.62% Weighted F1) and 61.47% of emotion accuracy (62.53% Weighted F1). The Dual-view SVM approach worked a bit worse, which means that there is not much effect of the character-level features added. Among the Transformer-based architectures, Twitter-XLM-RoBERTa got the best results on sentiment task with 71.90% accuracy and 71.98% Weighted F1 score thanks to the domain-specific pretraining on multilingual Twitter dataset. For emotion detection, the best performance was achieved by the XLM-R Multitask architecture with 64.97% accuracy and 64.26% Weighted F1 score.

International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 402-411 March 2026-Sept 2026 www.csjournals.com

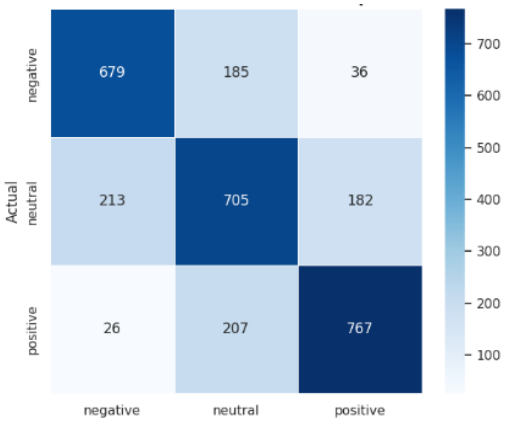


Fig 6: Confusion Matrix for Sentiment

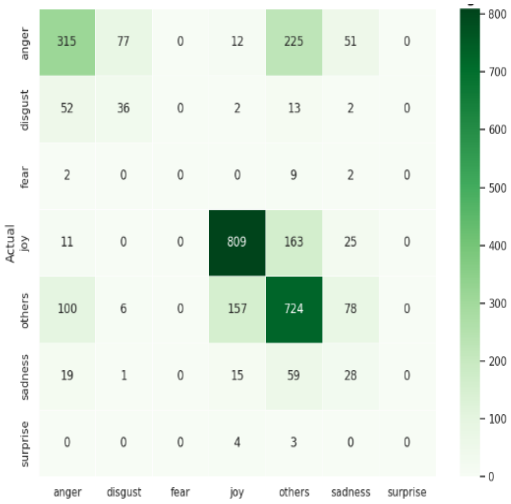


Fig 7: Confusion Matrix for Emotion

Model	Accuracy (%)	Precision (%)	Recall (%)	Weighted F1-score (%)
Dummy Baseline	35.50	18.60	35.50	18.60
Naïve Bayes	61.23	58.60	61.23	58.11
Logistic Regression	61.40	62.40	61.40	62.20
Linear SVM	61.47	65.00	61.47	62.53
Dual-view SVM	60.83	62.30	60.83	62.01
Distil-mBERT	**63.73**	**65.00**	**63.73**	**63.79**
mBERT	56.13	57.20	56.13	57.07
XLM-RoBERTa	57.37	59.50	57.37	59.22
Twitter-XLM-RoBERTa	58.93	60.10	58.93	59.62

Table 5: Classification Report for Sentiment

Model	Accuracy (%)	Precision (%)	Recall (%)	Weighted F1-score (%)
Dummy Baseline	36.67	19.67	36.67	19.67
Naïve Bayes	65.50	64.90	65.50	64.72
Logistic Regression	68.47	68.30	68.47	68.28
Linear SVM	68.97	69.00	68.97	68.62
Dual-view SVM	68.63	68.40	68.63	68.29
Distil-mBERT	69.57	69.60	69.57	69.39
mBERT	68.27	68.60	68.27	68.52
XLM-RoBERTa	69.90	69.90	69.90	69.75
Twitter-XLM-RoBERTa	**71.90**	**72.00**	**71.90**	**71.98**

Table 4: Classification Report for Emotion

International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 17 • Issue 2 pp. 402-411 March 2026-Sept 2026 www.csjournals.com

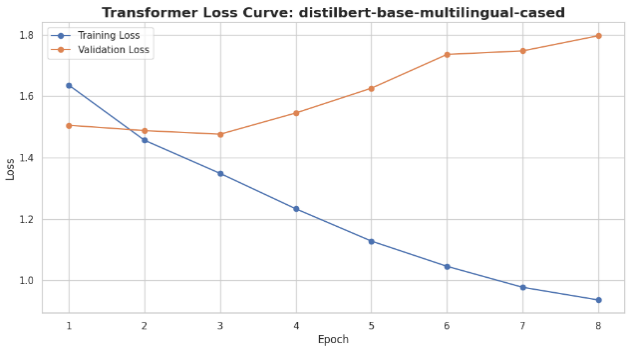


Fig 8: Transformer Loss Curve: distil-mBERT

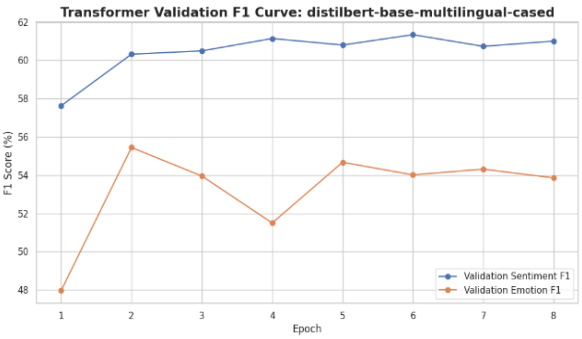


Fig 9: Transformer Validation F1 Curve

The training loss was continually decreasing throughout the epochs, which indicated successful learning. On the other hand, the minimum value of the validation loss was attained at epoch 3, after which there was an increase, indicating some degree of overfitting. The validation Sentiment F1-score was improving fast during the initial epochs, then remained constant, while the Emotion F1-score had low and unstable values due to class imbalance. These results imply that epoch 3 was the best epoch for generalization performance.

Demo - Predicting Both Sentiment & Emotion

Comment	Sentiment	Emotion	Sentiment Confidence	Emotion Confidence
I am so happy today, life is beautiful!	positive	joy	0.916	0.752
Yaar mujhe bahut gussa aa raha hai	neutral	others	0.870	0.692
This is really sad and depressing	negative	sadness	0.428	0.953
Great work by the team, well done!	positive	joy	0.918	0.782
Totally useless, waste of my time	negative	anger	0.858	0.496
Aaj toh bahut maza aaya yaar	positive	joy	0.896	0.783
Not sure how I feel about this	neutral	others	0.873	0.781
I love you so much, you mean everything	positive	joy	0.927	0.760
Bahut khushi ho rahi hai, finally mera kaam ho gaya!	positive	joy	0.905	0.776
Yaar mujhe bahut gussa aa raha hai, this is totally unfair	negative	sadness	0.554	0.885
This update is amazing, great work by the team!	positive	joy	0.921	0.772
Ye decision bahut bura hai, I completely hate it	negative	anger	0.894	0.488
I am not sure about this news, let's wait and see	neutral	others	0.891	0.759
This is disgusting and shameful behavior	negative	disgust	0.901	0.592

Fig 10: Demo Prediction for Sentiment & Emotion along with Confidence Weights

Key Observations

Ob 1: Sentiment was best performed by Twitter-XLM-RoBERTa, beating Linear SVM by 3.36% Weighted F1 due to domain-specific Twitter pretraining.

Ob 2: In case of emotion classification, Linear SVM stayed on par with Transformer architectures since bigger models need additional training and balance.

Ob 3: The best emotion results (64.26% Weighted F1) were scored by XLM-R Multitask using multitask learning, more training and class weights.

Ob 4: Fear and Surprise classes still proved challenging to classify due to very small number of training examples.

Ob 5: Convergence for sentiment happened at epoch 3, whereas emotion performance plateaued since then.

Conclusion

Thus, our research demonstrates that multilingual Transformer models perform better than the traditional ML algorithms for sentiment and emotion detection of code-switched Hinglish text. Twitter-XLM-RoBERTa reached the best results in sentiment classification task with Accuracy 71.90%, Weighted F1 71.98%, while XLM-R Multitask delivered the best scores in the emotion recognition task with Accuracy 64.97% and Weighted F1 64.26%. Our research indicates that representations learned by Transformer models work better with multilingual and code-mixed data. However, we also argue that class imbalance, especially with Fear and Surprise classes, is the main problem of emotion detection models.

Limitations

- There is a critical class imbalance problem for fear and surprise, resulting in very low recognition rates for these classes for all methods used.
- Romanised Hinglish is only considered; findings cannot be extended to other code-mixed language pairs such as Devanagari-script Hindi.
- This work only considers text data, no emoji usage, images, or thread conversation context is considered.
- Fine-tuning of large Transformers requires significantly more computing power than classical approaches.

Future Scope

- Future research could explore the use of data augmentation, SMOTE, focal loss, and class-balanced training for enhanced minority emotion detection.
- The architecture could be applied to other mixed Indian languages, including Tamil-English, Bengali-English, and Marathi-English.

- Using multimodal data (emojis, images, audio, and conversation) could increase classification accuracy.
- More advanced methods such as multitask Transformers, Large Language Models (LLMs), and RAG could be used for better sentiment analysis.

Acknowledgement

The author is highly thankful to Dr. Anupam Bhatia, Supervisor, DCSA, CRSU, Jind, for his immense support and constructive suggestions throughout this study. Highly thankful to Dr. Naveen, Mentor, for his immense help throughout the project. The author thanks Ms. Suman and Dr. Sneha for their ininvaluable advice & encouragement in this aforementioned study work. Lastly, the author is very thankful to the DCSA department, Chaudhary Ranbir Singh University, Jind, for providing the necessary infrastructure and environment for carrying out this research.

References & Bibliography

[1] S. Bhardwaj, *"SentiMix-EmoSen: Hinglish Sentiment and Emotion Dataset,"* Kaggle Dataset, 2025. [Online]. Available: https://www.kaggle.com/datasets/swatibhardwaj08/sentimix-emosen. [Accessed: 25-Jun-2026].

[2] A. Sharma and B. P. Singh, "ML Algorithms and Real World Application," *International Journal for Research in Engineering Application & Management (IJREAM)*, vol. 08, no. 11, pp. 2454–9150, 2023, doi: 10.35291/2454-9150.2023.0024.

[3] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision", Accessed: Apr. 18, 2026. [Online]. Available: http://tinyurl.com/cvvg9a.

[4] B. R. Chakravarthi, V. Muralidaran, R. Priyadharshini, and J. P. McCrae, "Corpus Creation for Sentiment Analysis in Code-Mixed Tamil-English Text," 2020. Accessed: Apr. 25, 2026. [Online]. Available: https://aclanthology.org/2020.sltu-1.28/.

[5] S. Ghosh, A. Priyankar, A. Ekbal, and P. Bhattacharyya, "Multitasking of sentiment detection and emotion recognition in code-mixed Hinglish data," *Knowl. Based. Syst.*, vol. 260, p. 110182, Jan. 2023, doi: 10.1016/J.KNOSYS.2022.110182.

[6] H. Zhu, "Sentiment analysis of 2021 Canadian election tweets," p. 9, Mar. 2023, doi: 10.1117/12.2667211.

[7] A. A. Nagra, K. Alissa, T. M. Ghazal, S. Saigeeta, M. M. Asif, and M. Fawad, "Deep Sentiments Analysis for Roman Urdu Dataset Using Faster Recurrent Convolutional Neural Network Model," *Applied Artificial Intelligence*, vol. 36, no. 1, 2022, doi: 10.1080/08839514.2022.2123094.

[8] F. Ullah, X. Chen, S. B. H. Shah, S. Mahfoudh, M. A. Hassan, and N. Saeed, "A Novel Approach for Emotion Detection and Sentiment Analysis for Low Resource Urdu Language Based on CNN-LSTM," *Electronics (Switzerland)*, vol. 11, no. 24, Dec. 2022, doi: 10.3390/electronics11244096.

[9] G. Singh, "Sentiment Analysis of Code-Mixed Social Media Text (Hinglish).

[10] A. Phadte and M. L. Dhore, "Sentiment Analysis of English-Marathi-Konkani Code-Mixed Social Media Text: A Multilingual Approach," in *2025 International Conference on Computing Technologies, ICOCT 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ICOCT64433.2025.11118344.

[11] A. M. D. Gowda, D. Vikram, and P. R. Hegde, "Bridging Linguistic Complexity: Sentiment Analysis of Tamil Code-Mixed Text Using Meta-Model," in *Proc. DravidianLangTech@NAACL*, 2025, pp. 103–108, doi: 10.18653/v1/2025.dravidianlangtech-1.18.

[12] A. M. D. Gowda, D. Vikram, and P. R. Hegde, "Bridging Linguistic Complexity: Sentiment Analysis of Tamil Code-Mixed Text Using Meta-Model," pp. 103–108, Jun. 2025, doi: 10.18653/V1/2025.DRAVIDIANLANGTECH-1.18.

[13] N. N. I. Prova, V. Ravi, M. P. Singh, V. K. Srivastava, S. Chippagiri, and A. P. Singh, "Multilingual sentiment analysis in e-commerce customer reviews using GPT and deep learning-based weighted-ensemble model," *International Journal of Cognitive Computing in Engineering*, vol. 7, no. 1, pp. 268–286, Dec. 2026, doi: 10.1016/J.IJCCE.2025.10.003.

[14] G. BANDARUPALLI, "Enhancing Sentiment Analysis in Multilingual Social Media Data Using Transformer-Based NLP Models A Synthetic Computational Study," Apr. 11, 2025. doi: 10.36227/techrxiv.174440282.23013172/v1.

[15] R. Singhal and S. Bishnoi, "HINGLISH SENTIMENT ANALYSIS," *International Advanced Research Journal in Science, Engineering and Technology (IARJSET)*, vol. 10, no. 1, 2023, doi: 10.17148/IARJSET.2023.10130.