International Journal of Engineering Research and Sustainable Technologies
Volume 3, No.1, March 2025, P 1- 9
ISSN: 2584-1394 (Online version)

COMPARATIVE ANALYSIS OF TRANSFORMER MODELS FOR SENTIMENT CLASSIFICATION IN CODE- MIXED INDIC LANGUAGES

[1]Mohana Priya K.T, [2]Shrinithi G, [3]Nithish P, [4]Panesh A C
[1]Assistant Professor, [1,2,3]UG Scholar
[1,2,3,4]Department of CSE, Kongu Engineering College Erode, India
*** Corresponding author email address**: mohanapriya.cse@kongu.ac.in

DOI: https://doi.org/10.63458/ijerst.v3i1.101 | ARK: https://n2t.net/ark:/61909/IJERST.v3i1.101

Abstract
Multiple language usage in a single message, or code-mixed text, has increased dramatically as a result of increased social media engagement. Because of this, activities involving Natural Language Processing (NLP), such as sentiment analysis and cyberbullying identification. Models that can effectively manage linguistic variability while retaining high accuracy are needed to address these issues. We investigate transformer-based designs that improve classification performance by utilizing knowledge transfer strategies. RoBERTa, GPT-2, XLM-RoBERTa, and IndicBERT are used in our method, which enhances classification accuracy by the transfer of sharing-private information across code-mixed and monolingual tasks. Results from experiments show that our multi-task framework surpasses single-task models with high accuracy on all datasets with:IndicBERT achieved 96.86% for Hinglish, XLM-RoBERTa achieved 96.95% for Punglish, and IndicBERT obtained 97.55% for Tanglish. In order to advance reliable NLP applications in multilingual environments, this project highlights the transformers' multi-task learning capabilities in enhancing performance on low-resource and code-mixed languages.

Keywords: Sentiment analysis, transformer-based models, code-mixed tasks, and indicator-BERT

1. Introduction

These days, social media sites like Twitter, YouTube, and Facebook are frequently used for communication and self-expression. As online contacts have increased, code-mixing—the practice of merging multiple languages in one single post—has become more wide spread. In multilingual nations like India, where people frequently mix English with their local tongues, this is particularly prevalent. Code-mixing facilitates more organic user expression, but it also presents difficulties for NLP applications like sentiment analysis and the detection of cyberbullying. Traditional language models are often trained on monolingual data, making it challenging to comprehend and handle such mixed-language content. Sentiment analysis is a crucial task in natural language processing by which sentiment or emotion behind a particular text is determined. Numerous applications, such as product reviews, social media monitoring, and customer feedback analysis, depend on it. Although significant progress has been made in sentiment analysis for monolingual languages such as English, the topic of code-mixed sentiment analysis remains relatively unexplored. Because code-mixed text frequently follows informal forms and lacks grammatical consistency, it is challenging for traditional NLP models to handle. Conventional sentiment analysis techniques use machine learning models trained on structured text data, rule-based techniques, and lexicons. Due to its unexpected nature, code-mixed text is difficult for these methods to handle. Deep learning and transformer-based models have become effective tools for multilingual and low-resource language challenges in recent years. Transformer-based designs have demonstrated encouraging results in comprehending linguistic patterns across several languages, including RoBERTa, GPT-2, XLM-RoBERTa, and IndicBERT. These models are ideal for processing code-mixed text since they are built to effectively extract contextual meanings.

In this work, we use transformer-based models to investigate the sentiment analysis of code-mixed datasets in Hinglish, Punglish, and Tanglish. Our method analyses and categorizes mixed-language content by utilizing pre-trained models. We hope to enhance sentiment categorization for code-mixed information by adding language representations and contextual embeddings. In multilingual contexts, transformer-based models are more successful at detecting sentiments than traditional approaches because they can manage intricate sentence structures and a wide range of vocabulary. Every day, social media sites produce enormous volumes of multilingual information, so it is crucial to create strong NLP models that can comprehend this type of writing. Efficient sentiment analysis of code-mixed content can improve user experience, assist companies in comprehending consumer viewpoints, and allow academics to conduct additional investigations of online debates. This study emphasizes how code-mixed NLP is becoming more and more significant and how sophisticated models are required to handle its particular difficulties. By looking into transformer-based solutions, we intend to contribute to the development of better, more accurate analysis of sentiment systems for multilingual and low-resource language environments. In order to improve language comprehension in the digital age, our work highlights the importance of bridging the gap between monolingual and multilingual NLP research.

2. Literature Review

The paper [1] predicts emotion, sentiment, and intensity, presenting a multi-task ensemble framework. It integrates CNNs, LSTMs, attention mechanisms, and traditional classifiers like SVMs and Random Forests. The model achieved 78.49% accuracy for sentiment classification and 74.16% for emotion classification, demonstrating innovative findings on data standards. The paper [2] suggests an approach with two phases that combines group learning and feature selection for aspect-based emotion analysis. It achieves 79.52% accuracy on benchmark datasets using models like as Naïve Bayes, SVM, and Decision Trees. The paper [3] offers an ensemble method for smooth monetary sentiment analysis based on multilayer perceptron (MLP).It integrates multiple classifiers including SVMs and decision trees, achieving an accuracy of 84.2% on benchmark financial datasets. The paper
[4] suggests using a mixed deep learning architecture to assess sentiment, combining CNNs and LSTMs to capture spatial and sequential features. The model achieves an accuracy of 86.2% on benchmark sentiment analysis datasets. The paper [5] explores Bangla text classification using BERT and its variants, transformer-based models. The proposed approach achieves an accuracy of 88.3% on benchmark Bangla text classification datasets. The paper [6] introduces a cross-lingual and multilingual approach for aspect-based sentiment analysis in order to handle data rarity. It utilizes bi-directional LSTMs with attention mechanisms, achieving an accuracy of 82.5% on multilingual sentiment datasets. The paper [7] uses connected WordNets to provide a cross-lingual sentiment analysis approach for Indian languages. Using SentiWordNet-based lexical translation and machine learning classifiers, it obtained 78.6% accuracy on benchmark datasets. The paper [8] offers a method for analysing the emotions of Bangla texts that combines an enlarged lexical dictionary with supervised machine learning models. Using classifiers like as Random Forest, SVM, and Naïve Bayes, it achieved an accuracy of 81.2% on benchmark datasets. The paper [9] proposes an aspect-based sentiment analysis approach using a fine-tuned BERT Base Uncased model. The model achieves an accuracy of 89.4% on benchmark sentiment analysis datasets, improving performance over traditional machine learning methods. The paper [10] proposes a domain-transferable lexical collection and supervised machine learning for Twitter emotion analysis. It employs classifiers like SVM and neural networks, achieving an accuracy of 85.3% on benchmark Twitter datasets.

The paper [11] uses machine learning alongside deep learning models to carry out sentiment analysis on Twitter data amidst India's COVID-19 lockdown. Using the gathered Twitter dataset, it obtained an accuracy of 87.6% by utilizing Naïve Bayes, SVM, and LSTM. The study [12] proposes an adversarial training approach for code-mixed sentiment classification. It utilizes a BiLSTM model with adversarial training and word embeddings, achieving an accuracy of 76.8% on the SemEval-2020 problem 9 dataset. The paper [13] presents an architecture for text classification using competitive learning with multiple tasks. It uses a shared-LSTM model with adversarial training to improve generalization across tasks, achieving an accuracy of 90.3% on benchmark text classification datasets. The study in [14] focuses on building and assessing emotion analysis of multi-domain tweet corpora. It employs deep learning and machine learning models, such as LSTMs and BERT, and on the recently generated dataset, it achieved an accuracy of 87.5%. The paper [15] employs Support Vector Machines (SVM), Maximum Entropy, and Naïve Bayes for sentiment classification. SVM produced the highest accuracy of 82.9% on a dataset of movie reviews. The paper [16] presents HindiMD, a multi-domain corpus for sentiment evaluation in Hindi with few resources. It employs LSTMs, CNNs, and transformer-based models like BERT, achieving an accuracy of 85.2% on the benchmark dataset. The paper [17] gives a summary of SemEval-2020 Task 9, which was concerned with sentiment analysis of tweets that were code-mixed in Hindi and Spanish. A range of models were employed, encompassing deep learning models (LSTMs, CNNs, and transformer-based models like BERT) and conventional machine learning techniques (SVM, Naïve Bayes). Top-performing systems reached the highest accuracy of almost 73% for the Hindi-English dataset and 75% for the Spanish-English dataset.

The paper [18] explores the capabilities of Multilingual BERT (mBERT) in handling cross-lingual transfer learning for various languages. Zero-shot cross-lingual performance is assessed on tasks such as dependent parsing and part-of-speech tagging. The reported accuracy varies by task and language, but mBERT demonstrates strong performance, often achieving over 90% accuracy on POS tagging for high-resource languages. The paper [19] evaluates Naïve Bayes (NB) and Support Vector Machine (SVM) for sentiment classification. The findings demonstrate that SVM performs better than Naïve Bayes, with an accuracy of 85.29% as opposed to 81.12% for Naïve Bayes. The study emphasizes that SVM's superior performance is a result of its competent handling of feature spaces with large dimensions. The paper [20] uses Support Vector Machine (SVM) and Naïve Bayes for sentiment analysis of code-mixed text. The highest accuracy achieved was 78.5% using SVM, which outperformed

Naïve Bayes. The study also focuses on text normalization techniques to enhance code-mixed data emotional classification effectiveness.

3. Methodology

Transformer models, which are renowned for achieving cutting-edge results on text categorization problems, are examined in this work. Among the models utilized are GPT-2, RoBERTa, XLM-RoBERTa, and IndicBERT. The classifiers utilize token embeddings generated by these models to categorize tweets into two predetermined groups: "positive" and "negative". Key aspects extracted from the dataset include tokenized text, contextual word relationships, and semantic embeddings. Additional preliminary processing methods, including stop-word removal and removing user mentions (@username), are used to improve feature quality because they ultimately have little impact on sentiment prediction. Figure 1 below shows the workflow of the study, and the following sections provide a detailed explanation of the framework of the proposed model. The following significant topics are discussed in this section: (1) collecting the dataset (2) preprocessing it (3) Dataset augmentation (4) implementing GPT-2,(5)deploying RoBERTa, XLM-RoBERTa, and IndicBERT, and (6) assessing performance using relevant metrics are the subsequent actions.

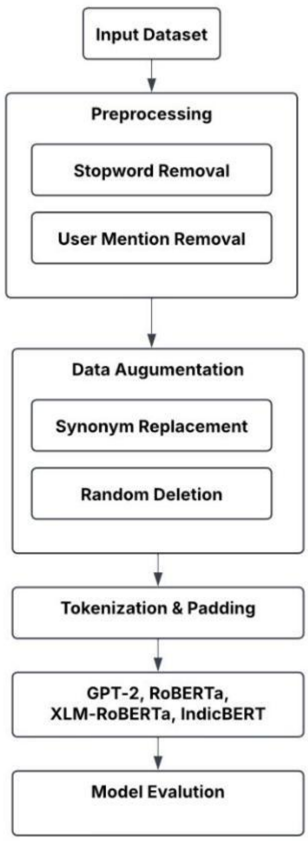


Fig 1. Proposed System Architecture

The suggested system flow starts with data preprocessing, where tweets are polished by eliminating special characters, superfluous phrases, and extraneous content, as shown in Figure 1. Tokenization is then used to standardize the format of the text. Transformer-based models like GPT-2, RoBERTa, XLM-RoBERTa, and IndicBERT are then used to train these tagged tweets in order to predict whether or not they are cyberbullying. The proposed model makes use of the transformer architecture, which tokenizes input text x into subwords. Each token is part of the high-dimensional vector e(x). The attention mechanism assigns a weight to each token based on its contextual relevance, as demonstrated below:

$$(\, , \, , \,) = \quad (\quad \sqrt{dk}) \tag{1}$$

In this case, d_k is the dimension of the keys, while Q, K, and V are the query, key, and value matrices. To anticipate the target class, the output embeddings are further run through a classification layer.

a. Data Collection

Sentiment analysis of multilingual text has been done using the hate-speech dataset from social media sites. 18,148

Hinglish (Hindi+English) text samples that have been labeled for sentiment classification make up the Hinglish dataset. 3,070 Punglish (Punjabi+English) text samples that have been labeled for sentiment classification are included in the Punglish dataset. 990 Tanglish (Tamil+English) text samples that have been similarly categorized for sentiment classification are included in the Tanglish dataset. In order to examine multilingual sentiment detection in code-mixed languages, these datasets have been assembled.

b. Data Preprocessing

The data is subjected to several preprocessing steps to clean and refine the text before sentiment classification. To filter out common keywords that contribute little to sentiment analysis, stopwords were removed. To avoid links interfering with text-based classification, URLs were stripped off. Names (split by '@') were also removed to avoid user-specific words from affecting sentiment detection. Further, the dataset was enriched and model generalization was achieved with data augmentation strategies. Synonym substitution introduced tweaks while keeping the original sense intact by replacing words with their synonyms. The preprocessing and augmentation procedures were methodically performed to the Hinglish, Punglish, and Tanglish datasets in order to improve the classification algorithm's precision and resilience.

c. Feature Extraction

Token embeddings were extracted using transformer-based models, notably GPT-2, RoBERTa, XLM-RoBERTa, and IndicBERT. The text was tokenized into subword units during preprocessing, and these previously trained models were used to create embeddings. These embeddings served as the classifier's input features, capturing contextual and semantic information.

d. Model Selection

Transformers are better suited to catch contextual dependencies and subtleties in text data. GPT-2 was used due to its high autoregressive strengths that allow it to create strong contextual embeddings. RoBERTa was used because of its efficient pretraining strategies and dynamic masking that help to increase model generalization. XLM-RoBERTa was added due to its multilinguality, so it was perfect for code-mixed language handling. IndicBERT was chosen for its proficiency in processing Indian languages, leading to a better context-sensitive representation of Hinglish, Punglish, and Tanglish text.

e. *GPT-2 Base Model*

GPT-2, one of the strong autoregressive NLP transformer models, was employed in the present study. GPT-2 is suited to represent the contextual sequences of text because it produces text left to right and not like the bidirectional model. The few transformer layers involving self-attention in the model enable it to capture intricate dependencies between words. The standard model of GPT-2 uses 117 million parameters, 12 decoder layers, and 12 attention heads. It excels in downstream tasks like sentiment estimation, text categorization, and language modeling because it has previously been prepared on a vast dataset of web content of various kinds. In addition, it is better at sentiment classification tasks when fine-tuning on labelled data because it enhances its contextual understanding.

f. XLM-RoBERTaModel

Since it supports processing text in so many languages, the multilingual RoBERTa variant XLM-RoBERTa, or XLM-R, is a good choice for code-mixed language work. It removes Next Sentence Prediction (NSP), which was an element of BERT, and applies dynamic masking to improve generalization of the model.With training in more than 100 languages, XLM-R is able to identify complex linguistic links and structures in text that has been combined with code. XLM-R is suitable for sentiment classification tasks on Hinglish, Punglish, and Tanglish datasets due to its exceptional cross-lingual capacity. On NLP tasks, XLM-R can outperform the majority of traditional multilingual models thanks to its incredibly big dataset and creative training methods.

g. RoBERTa Model

For improved performance, pretraining techniques are optimized, similar to RoBERTa, a BERT version. RoBERTa significantly improves contextual word representations when the Next Sentence Prediction (NSP) objective is removed and dynamic masking and higher batch sizes are used. Because RoBERTa can detect extremely fine syntactic and semantic links, it can be used for sentiment classification and code-mixed text analysis. RoBERTa was demonstrated to be a competitive model in both classification and language-model tasks, outperforming BERT in numerous NLP examinations.

h. IndicBERT Model

IndicBERT is a transformer model created specifically for Indian languages. Unlike universal multilingual models, IndicBERT was trained on a multi-Indic language dataset, which makes it extremely successful for code-mixed text in Hinglish, Punglish, and Tanglish. By gathering both left and right context through bidirectional attention mechanisms, it enhances word representations. Its efficiency and linguistic diversity make it a successful model for sentiment classification in code-mixed language and low-resource contexts.

i. Performance Evaluation

The performance of the classification models (GPT-2, RoBERTa, XLM-RoBERTa, and IndicBERT) is evaluated using accuracy as the primary metric. The degree to which a model can correctly classify examples by evaluating whether its predictions are correct is known as accuracy. It only verifies the model's projected label's accuracy out of the entire set. The formula for accuracy is given as:

$$= \frac{+}{+ + +} \tag{2}$$

Equation (1) refers to the accuracy formula where TP= True Positive, TN= True Negative, FP= False Positive and FN=False Negative.

A tabular representation that displays the count of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each class is called the confusion matrix. In the case of binary classification, it is structured as a 2×2 matrix, summarizing the model's performance by indicating how many instances were correctly or incorrectly classified.

Table 1. Confusion Matrix

	Predicted Positive (PP)	**Predicted Negative (PN)**
Actual Positive (AP)	True Positive (TP)	False Negative (FN)
Actual Negative (AN)	False Positive (FP)	True Negative (TN)

4. Result and Discussions

This section illustrates how well the transformer-based models GPT-2, RoBERTa, XLM-RoBERTa, and IndicBERT perform sentiment classification on a social media dataset. With a batch count of 16 and a learning rate of 2e-5, the models were trained over ten epochs. The dataset was divided 80:20 between the training and testing halves.The models stabilized and produced the best performance outcomes after the tenth session. The results of accuracy-based testing are shown in the following table.

Table 2. Results of Proposed Models

	Hinglish	**Punglish**	**Tanglish**
GPT_2	0.953500	0.955500	0.92700
RoBERTa	0.958667	0.958500	0.968300
XLM_RoBERTa	0.959667	0.96400	0.976500
IndicBERT	0.962667	0.957000	0.978000

Performance was optimized by aggressive preprocessing techniques, including username removal and stop-word removal, which minimized noise in the data. Dataset augmentation methods also assisted in enhancing model generalization, while transformer models such as GPT-2, RoBERTa, XLM-RoBERTa, and IndicBERT provided high contextual accuracy. Performance metrics were also improved via hyperparameter tweaking, which included adjusting the number of epochs, learning rate, and dropout rate. For increasing accuracy, augmentation methods were utilized to strengthen data diversity as well as support balanced class balances. This kept sentiment categories adequately represented, with resulting classification output becoming stronger as well as unprejudiced. GPT-2, The GPT-2 model performed well, with accuracy measures of 95.35% for Hinglish, 95.55% for Punglish, and 96.27% for Tanglish. Although GPT-2 has a reputation for its ability to generate text, its capacity to understand deeper contextual relationships makes it highly effective in sentiment classification. Yet, the fact that it had slightly lesser accuracy than other transformer-based models indicates that autoregressive models may not be as suitable for classification as bidirectional models. RoBERTa , RoBERTa demonstrated better accuracy compared to GPT-2, with 95.87% for Hinglish, 95.85% for Punglish, and 96.83% for Tanglish. RoBERTa's optimization techniques, including dynamic masking and removal of the Next Sentence Prediction task, enhance its ability to understand complex linguistic patterns. Its performance suggests that the model effectively learns contextual dependencies, making it well-suited for sentiment classification in code-mixed languages. XLM-RoBERTa , XLM-RoBERTa surpassed GPT-2 and RoBERTa, scoring 95.97% on Hinglish, 96.40% on Punglish, and 97.65% on Tanglish. The model, which is built for multilingual and cross-lingual NLP tasks, performs well for code-mixed languages. In conclusion, XLM-RoBERTa is a good option for sentiment analysis in code-mixed contexts because of its utilization of multilingual embeddings, which enhances its classification accuracy.

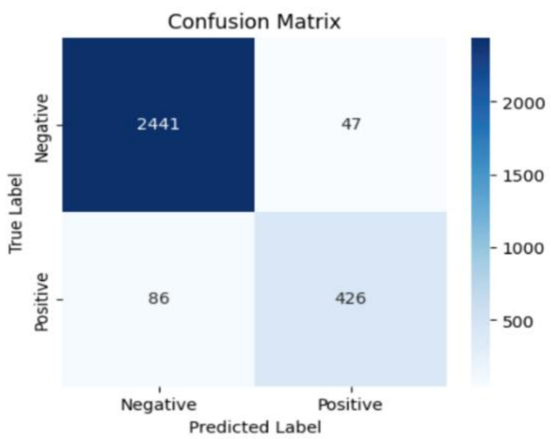


Fig. 2 Confusion Matrix for IndicBERT-Hinglish

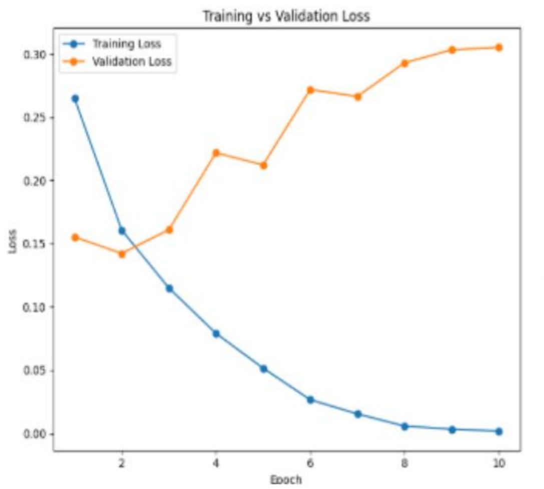


Fig.3 Accuracy graph for IndicBERT-Hinglish

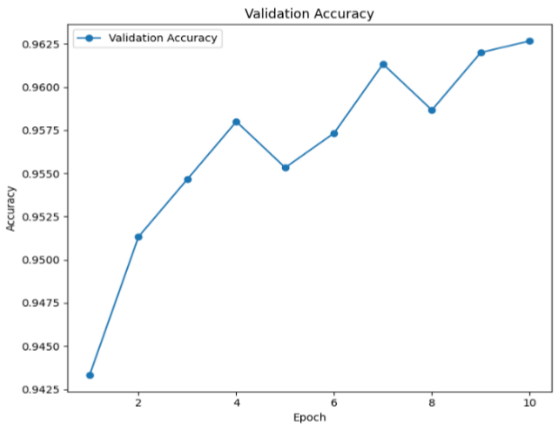


Fig 4 Loss graph for IndicBERT-Hinglish

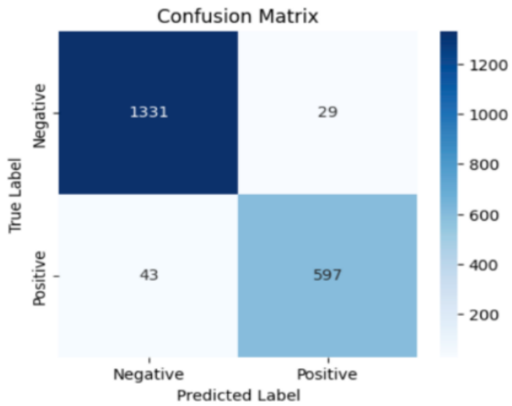
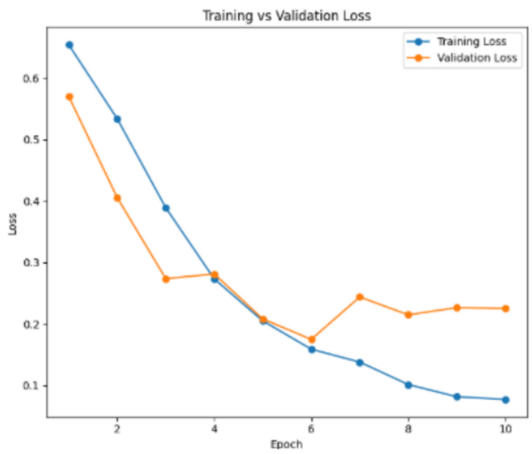


Fig 5 Confusion Matrix for XLM-RoBERT-Punglish

IndicBERT, surpassing all other corpora, IndicBERT reported 96.27% accuracy on Hinglish, 95.70% accuracy on Punglish, and 97.80% accuracy on Tanglish. Being the most efficient model to perform sentiment analysis in Hinglish, Punglish, and Tanglish, IndicBERT owes its success to its pre-trained model for Indian languages. As such, with its high performance, pretraining within the domain is crucial for higher accuracy in code-mixed and low-resource languages.

IndicBERT worked exceptionally well for Hinglish and Tanglish, proving the benefit of domain-specific pretraining for Indian languages. XLM-RoBERTa performed the best for Punglish, proving its capability to deal with multilingual and cross-lingual text efficiently. RoBERTa and GPT-2, though producing competitive results, were slightly less efficient, indicating that their architectures might not be optimized to the extent for the intricacies of code-mixed sentiment classification.

Overall, these results establish the performance of transformer-based models for addressing sentiment analysis in low-resource and multilingual environments, with IndicBERT being the optimal model for Hinglish and Tanglish, and XLM-RoBERTa for Punglish. Ensemble learning strategies, more hyperparameter fine-tuning, and expanding the dataset's size to enable generalization and stability across even more linguistic variants could be a subject of future work.



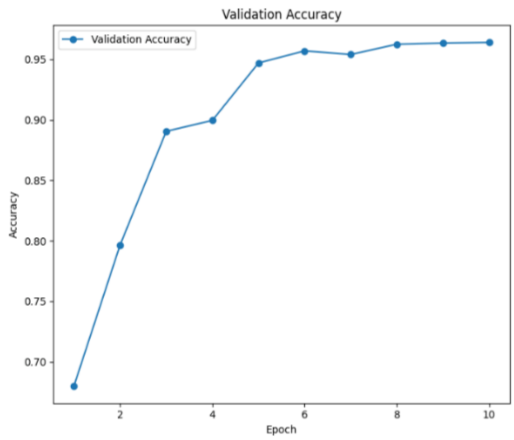


Fig. 6 Accuracy graph for XLM-RoBERT-Punglish **Fig.7** Loss graph for XLM-RoBERT-Punglish

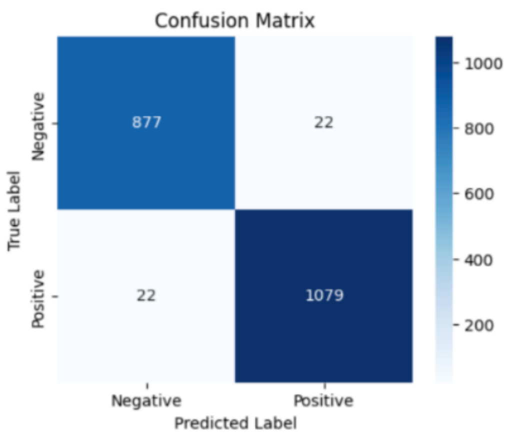


Fig. 8 Confusion Matrix for IndicBERT-Tanglish **Fig. 9** Accuracy graph for IndicBERT-Tanglish

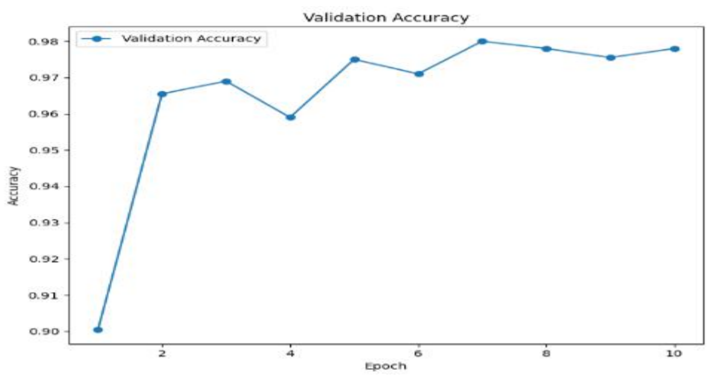


Fig.10 Loss graph for IndicBERT-Tanglish

5. Conclusion

This study shows how transformer-based models are effective for sentiment classification in code-mixed languages like Hinglish, Punglish, and Tanglish. Our multi-task learning approach improves accuracy by transferring knowledge between code-mixed and monolingual tasks. Experimental data show that IndicBERT outperforms RoBERTa, GPT-2, and XLM-RoBERTa on Tanglish. Various transformer topologies (such as T5, mT5), fine-tuning on bigger datasets, and incorporating outside language resources can all be explored in future work. Contextual knowledge can be further enhanced by contrastive or reinforcement learning, and real-world application will be improved by branching out to low-resource languages.

References

1. Akhtar, M.S., Ghosal, D., Ekbal, A., et al., 'A multi-task ensemble framework for emotion, sentiment and intensity prediction'. https://doi.org/10.48550/arXiv.1808.01216. 2018
2. Akhtar, M. S., Gupta, D., Ekbal, A., et al., 'Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis'. Knowledge-Based Systems, 125, 116–135. https://doi.org/ 10.1016/j.knosys.2017.03.020. 2017
3. Akhtar, M.S., Kumar, A., Ghosal, D., et al., 'A multilayer perceptron based ensemble technique for fine-grained financial sentiment analysis'. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 540–546. Association for Computational Linguistics. https://doi.org/ 10.18653/v1/D17-1057. 2017
4. Akhtar, M.S., Kumar, A., Ekbal, A., et al., 'A hybrid deep learning architecture for sentiment analysis. In: Proceedings of COLING 2016', the 26th International Conference on Computational Linguistics: Technical Papers. 482–493. The COLING 2016 Organizing Committee. https://aclanthology.org/C16- 1047. 2016
5. Alam, F., 'Bangla text classification using transformers'. https://doi.org/10.48550/ arXiv.2011.04446.2020
6. Akhtar, M.S., Sawant, P., Sen, S., et al., 'Solving data sparsity for aspect based sentiment analysis using cross-linguality and multi-linguality'. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). 572–582. Association for Computational Linguistic. https://doi.org/10.18653/v1/N18- 1053.2018
7. Balamurali, A., Joshi, A., Bhattacharyya, P., 'Cross-lingual sentiment analysis for indian languages using linked wordnets. In: Proceedings of COLING 2012: Posters. 73–82. The COLING 2012 OrganizingCommittee. https://aclanthology.org/C12-2008. 2012
8. Bhowmik, N. R., Arifuzzaman, M., Mondal, M. R. H., et al., 'Bangla text sentiment analysis using supervised machine learning with extended lexicon dictionary'. Natural Language Processing Research, 1(3–4), 34–45. https://doi.org/10.2991/nlpr.d.210316.001.2021
9. Geetha, M., & Renuka, D. K., 'Improving the performance of aspect based sentiment analysis using fine-tuned bert base uncased model'. International Journal of Intelligent Networks, 2, 64–69. https://doi. org/10.1016/j.ijin.2021.06.005.2021
10. Ghiassi, M., & Lee, S. ,'A domain transferable lexicon set for twitter sentiment analysis using a supervised machine learning approach'. Expert Systems with Applications, 106, 197–216. https://doi.org/10. 1016/j.eswa.2018.04.006.2018
11. Gupta, P., Kumar, S., Suman, R., et al., 'Sentiment analysis of lockdown in india during covid-19: A case study on twitter'. IEEE Transactions on Computational Social Systems. https://doi.org/10.1109/TCSS. 2020.3042446.2020
12. Liu, J., Chen, X., Feng, S., et al., 'Kk2018 at SemEval-2020 task 9: Adversarial training for code-mixing sentiment classification'. In: Proceedings of the Fourteenth Workshop on Semantic Evaluation. 817–823. International Committee for Computational Linguistics, Barcelona (online). https://doi.org/10.18653/ v1/2020.semeval-1.103.2020
13. Liu, P., Qiu, X., Huang, X., 'Adversarial multi-task learning for text classification'. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 1–10. Association for Computational Linguistics, Vancouver, Canada. https://doi.org/10.18653/v1/P17- 1001.2017
14. Mamta, M., Ekbal, A., Bhattacharyya, P., et al., 'Multi-domain tweet corpora for sentiment analysis: Resource creation and evaluation'. In: Proceedings of the 12th Language Resources and Evaluation Conference. 5046–5054. European Language Resources Association, Marseille, France, https://aclanthology. org/2020.lrec-1.621.2020
15. Pang, B., Lee, L., Vaithyanathan, S., 'Thumbs up?: sentiment classification using machine learning techniques'. In: Proceedings of the ACL-02 conference on Empirical methods in natural language

processing-Volume 10. 79–86. Association for Computational Linguistics. https://doi.org/10.3115/ 1118693.1118704. 2002
16. Mamta, M., Ekbal, A., Bhattacharyya, P., et al., 'HindiMD: A multi-domain corpora for low-resource sentiment analysis. In: Proceedings of the Thirteenth Language Resources and Evaluation Conference. 7061–7070. European Language Resources Association, Marseille, France. https://aclanthology.org/ 2022.lrec-1.764.2022
17. Patwa, P., Aguilar, G., Kar, S., et al., 'SemEval-2020 task 9: Overview of sentiment analysis of code-mixed tweets'. In: Proceedings of the Fourteenth Workshop on Semantic Evaluation. 774–790. International Committee for Computational Linguistics, Barcelona (online). https://doi.org/10.18653/ v1/2020.semeval-1.100. 2020
18. Pires, T., Schlinger, E., Garrette, D., 'How multilingual is multilingual BERT?' In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 4996–5001. Association for Computational Linguistics, Florence, Italy. https://doi.org/10.18653/v1/P19-1493. 2019
19. Rana, S., Singh, A., 'Comparative analysis of sentiment orientation using svm and naive bayes techniques'. In: 2016 2nd International Conference on Next Generation Computing Technologies (NGCT).106–111.IEEE. https://doi.org/10.1109/NGCT.2016.7877399. 2016
20. Sharma, S., Srinivas, P., Balabantaray, R.C., 'Text normalization of code mix and sentiment analysis'. In: 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI). 1468– 1473. IEEE. https://doi.org/10.1109/ICACCI.2015.72758 , 2015