Tri-Lingual Sentiment Detection of Code-Mixed Marathi, Hindi, and English through a Hybrid Multilingual Model

Dr. Aniruddha Shelotkar[1], Dr. Sai Kiran Oruganti[2]

[1]Post-Doctoral Researcher, Computer Science Engineering, Lincoln University College, Malaysia (LUC/CPGS/PDF/2025150801); [2]Associate Professor, Lincoln University College, Malaysia

Email: [1]pdf.Aniruddha@lincoln.edu.my, [2]saisharma@lincoln.edu.my

Abstract—Sentiment analysis (SA) of code-mixed, multi-script text remains difficult for classical and transformer-based models alike, particularly in India, where Marathi, Hindi, and English are routinely mixed within a single sentence or conversation. This paper proposes a hybrid sentiment analysis model for tri-lingual code-mixed text that combines Multilingual BERT (mBERT) embeddings with an entropy-guided, script-aware attention mechanism and three language-specific experts (Marathi, Hindi, English), whose outputs are combined in a feature fusion layer for final sentiment classification. The attention mechanism dynamically re-weights token contributions using an entropy measure of language-switching uncertainty, allowing the model to adapt to code-mixing within a sentence rather than treating it as noise. The model is evaluated on a newly curated corpus of 50,000 code-mixed Marathi–Hindi–English samples drawn from social media, customer reviews, and chat data, labeled positive, negative, or neutral, and is compared against mBERT, MuRIL, and XLM-R baselines. The proposed hybrid model achieves 87.3% accuracy, 86.1% precision, 88.2% recall, and an 87.1% F1-score, outperforming the strongest baseline (XLM-R) by 5.8 accuracy points and the mBERT baseline by 11.9 points. These results indicate that combining script-aware, entropy-guided attention with language-specific experts is an effective strategy for tri-lingual code-mixed sentiment analysis, with direct applications to social media monitoring, multilingual customer feedback analysis, and sentiment-driven decision-making in code-mixing-heavy environments such as India.

Keywords—tri-lingual sentiment analysis; code-mixed text; Multilingual BERT (mBERT); hybrid model; script-aware attention; entropy-guided attention; Marathi; Hindi

I. INTRODUCTION

Sentiment analysis (SA) categorizes the emotion expressed in text and is a core task in natural language processing. Classical SA models were developed for monolingual text, where classification is comparatively straightforward. The growth of social media has made communication increasingly multilingual and code-mixed, particularly in India, where speakers routinely switch between Marathi, Hindi, and English within a sentence or conversation. This code-switching, combined with the use of two scripts (Devanagari for Marathi/Hindi, Latin for English), introduces syntactic variation, tokenization difficulty, and culturally specific sentiment expression that classical SA models and even multilingual transformers such as mBERT struggle to handle [2], [8]. More broadly, cross-lingual and cross-script text classification is known to be harder than monolingual classification even without code-mixing [12], which compounds when script-switching occurs within a single sentence.

Prior work on code-mixed sentiment analysis has concentrated mainly on bilingual Hindi–English settings — for example, the widely used Hindi–English sub-word-level sentiment corpus and model of Joshi et al. [6] and the SemEval-2020 Task 9 (SentiMix) shared task on Hindi–English and Spanish–English code-mixed sentiment [7] — while the tri-lingual case involving Marathi is comparatively unstudied. Existing multilingual models such as MuRIL [3] and IndicBERT [4] were not designed with explicit mechanisms for three-way script-aware code-switching, and broader surveys of code-switched language processing [8] similarly identify script- and language-aware modeling as an open problem. This motivates the research question addressed here: how can a hybrid sentiment analysis model be built to reliably classify sentiment in text that freely mixes Marathi, Hindi, and English at the word, phrase, and sentence level?

Contributions

- A hybrid architecture combining mBERT embeddings with a novel code-switching-sensitive, entropy-guided script-aware attention mechanism and three language-specific experts (Marathi, Hindi, English), the latter inspired by the mixture-of-experts paradigm [11].
- An entropy-guided attention formulation that dynamically re-weights token attention according to the model's uncertainty over the active language/script, rather than treating script changes as noise.
- Evaluation on a newly curated 50,000-sample tri-lingual code-mixed dataset, benchmarked against mBERT, MuRIL, and XLM-R.
- Interpretability via attention-map and entropy-value visualization, giving qualitative insight into the model's decisions on code-mixed input.

The remainder of the paper is organized as follows: Section 2 reviews related work; Section 3 presents the proposed methodology, dataset, architecture, and attention formulation; Section 4 describes the experimental setup and evaluation metrics; Section 5 reports and discusses results; and Section 6 concludes with future work.

II. Related Work

Table 1 summarizes prior work on multilingual and code-mixed sentiment analysis. mBERT [2] generalizes across languages via shared subword representations but degrades on code-mixed, multi-script input. MuRIL [3] and IndicNLPSuite/IndicBERT [4] improve on Indian-language coverage but target bilingual or monolingual settings with no explicit script-switching mechanism. XLM-R [5] offers a strong general-purpose cross-lingual baseline but likewise lacks script-aware attention. Task-specific Hindi–English work — Joshi et al. [6] and the SemEval-2020 SentiMix shared task [7] — established strong bilingual baselines, and the code-switching survey of Sitaram et al. [8] and the GLUECoS benchmark [9] identify script- and language-identification as open challenges, but none addresses the three-way Marathi–Hindi–English case or a Marathi corpus such as L3Cube-MahaSent [10]. This gap — a tri-lingual, script-aware, entropy-guided hybrid architecture with dedicated per-language experts — is the specific contribution of the present work.

Table 1. Comparative Summary of Related Work

Study	Focus	Key Contribution	Limitation Relative to This Work
Pires et al. (2019) [2]	mBERT for multilingual tasks	Demonstrated mBERT's cross-lingual generalization ability	Struggles with code-mixed, multi-script data
Khanuja et al. (2021) — MuRIL [3]	Multilingual model for Indian languages	Trained on transliteration/translation pairs for Indian languages	Targets bilingual Hindi–English; no Marathi or tri-lingual support
Kakwani et al. (2020) — IndicNLPSuite/IndicBERT [4]	Pre-trained resources for 11 Indian languages	Large monolingual corpora, embeddings, and IndicBERT model	No script-aware mechanism for in-sentence code-switching
Conneau et al. (2020) — XLM-R [5]	Large-scale cross-lingual pretraining	Strong general-purpose multilingual baseline across 100 languages	No script-aware or entropy-guided mechanism for code-switching
Joshi et al. (2016) [6]	Hindi–English code-mixed sentiment	Sub-word-level model and annotated Hi-En corpus	Bilingual only; does not address Marathi or tri-lingual mixing
Patwa et al. (2020) — SemEval-2020 Task 9 [7]	Shared task on code-mixed sentiment (Hi-En, Es-En)	Released annotated Hinglish/Spanglish corpora and benchmarked systems	Bilingual pairs only; no Marathi or three-way code-switching
Sitaram et al. (2019) [8]	Survey of code-switched speech/NLP	Comprehensive review of datasets, tasks, and open challenges	Does not propose a script-aware attention architecture
Khanuja et al. (2020) — GLUECoS [9]	Evaluation benchmark for code-switched NLP	Standardized benchmark across multiple code-switched tasks	Benchmark/evaluation focus; not a tri-lingual sentiment model

III. Methodology

This section describes the dataset, the hybrid model architecture, and the entropy-guided script-aware attention mechanism used to process code-mixed Marathi–Hindi–English text. Fig. 1 shows the overall model concept: mBERT

embeddings are passed through script-aware attention, routed to three language-specific experts, and combined in a feature fusion layer for final sentiment classification.

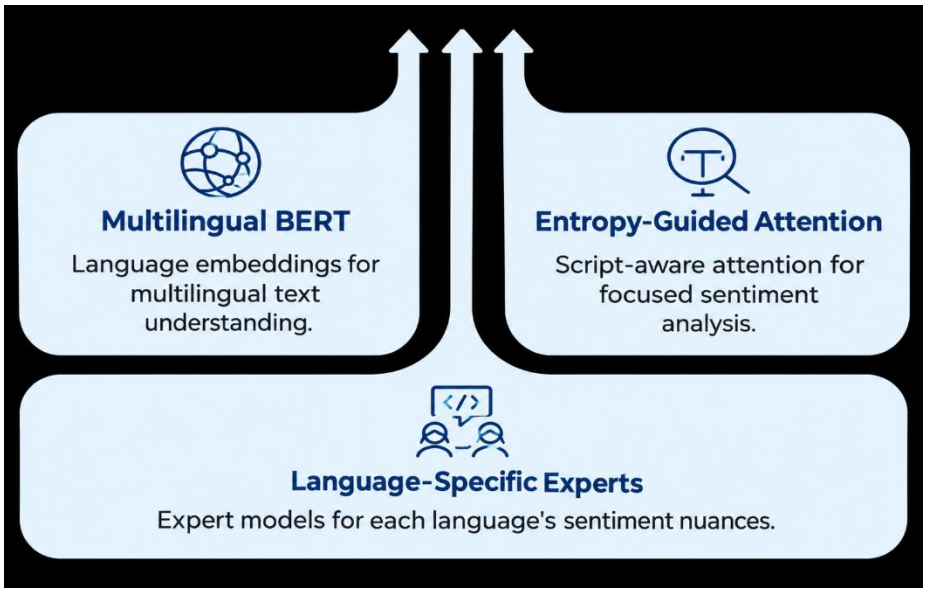

Figure 1. Conceptual hybrid model for tri-lingual sentiment analysis.

a. Dataset

The dataset comprises 50,000 code-mixed samples of Marathi, Hindi, and English collected from social media, customer reviews, and chat applications, cleaned of special characters and noise, and labeled positive, negative, or neutral. Preprocessing includes mBERT multilingual tokenization [2] across Devanagari and Latin scripts, language identification per token, and data augmentation that simulates sentence- and token-level code-switching to increase coverage of switching patterns. The dataset is split 80% training / 10% validation / 10% test.

b. Hybrid Model Architecture

The model has four components: (i) mBERT, which generates contextual token embeddings across the three languages [2]; (ii) an entropy-guided script-aware attention layer; (iii) three language-specific experts (Marathi, Hindi, English), each a neural sub-component specialized in the syntactic and semantic patterns of its language and conceptually related to the mixture-of-experts paradigm [11]; and (iv) a feature fusion layer that merges expert outputs into a single representation for sentiment classification. Fig. 2 shows the detailed architecture.

c. Entropy-Guided Script-Aware Attention

Let the mBERT token sequence be $T = [t_1, t_2, \ldots, t_n]$, with each token t_i mapped to an embedding $e_i \in \mathbb{R}^d$. Building on the general attention formulation of [1], [13], the attention weight for token i with respect to language context c (Marathi, Hindi, or English) is:

$$\alpha_i = \exp(\text{score}(e_i, c)) / \Sigma_j \exp(\text{score}(e_j, c)) \quad (1)$$

where $\text{score}(e_i, c)$ measures the relevance of token t_i to context c. The entropy of the resulting attention distribution,

$$H(\alpha) = -\Sigma_i \alpha_i \log(\alpha_i) \quad (2)$$

quantifies the model's uncertainty over language-switching at each position, and is used to guide re-weighting of attention as script/language confidence varies across the sentence — higher entropy signals a likely code-switch point requiring stronger reliance on the corresponding language expert. This entropy-guided formulation is, to our knowledge, a novel contribution of this paper rather than an adaptation of a prior published mechanism.

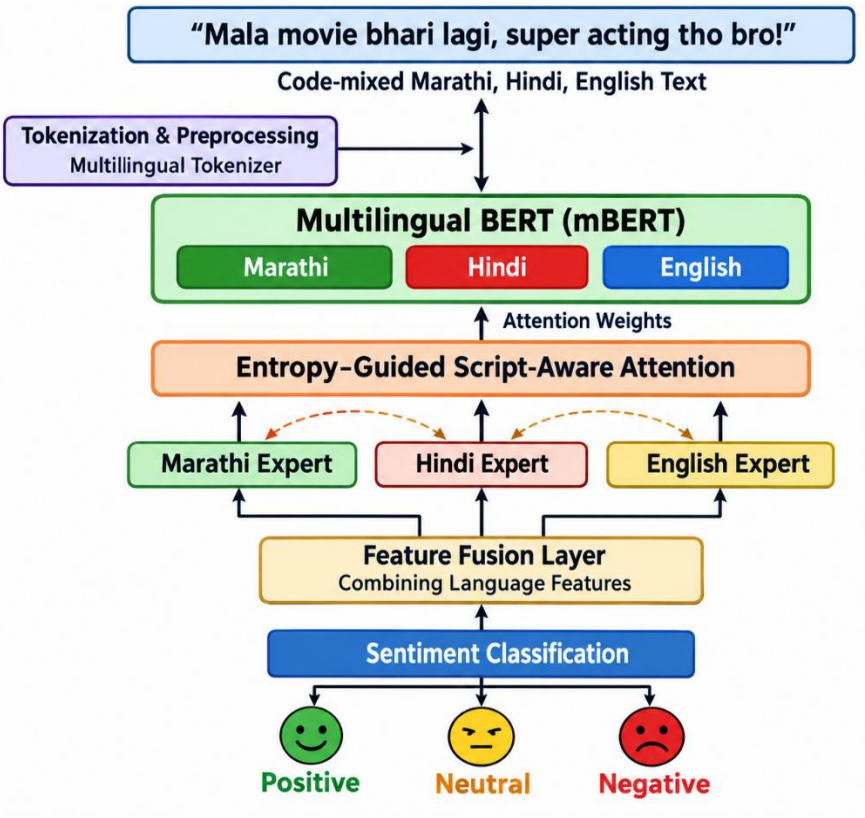

Figure 2. Detailed architecture of the proposed hybrid sentiment analysis model.

d. Feature Fusion and Classification

The outputs of the three language experts, F_i for i ∈ {Marathi, Hindi, English}, are combined via a weighted sum:

$$F = \Sigma_{i=1}^{3} \lambda_i F_i \quad (3)$$

where λ_i is the attention weight assigned to language i. The fused vector F is passed through a fully connected layer with softmax output over {positive, neutral, negative}.

e. Training and Optimization

The model is trained with a multi-task objective in the spirit of [14]: the primary task is sentiment classification, and the secondary task is language-switching detection, which regularizes the model toward script-aware representations. The total loss is:

$$L\_total = \lambda_1 L\_sentiment + \lambda_2 L\_switching \quad (4)$$

where L_sentiment and L_switching are cross-entropy losses for sentiment classification and language-switching detection, respectively, and λ_1, λ_2 are task weights. The model is optimized with Adam [15] at a learning rate of 1e-5, fine-tuned to reduce overfitting on code-mixed data.

IV. Experimental Setup

a. Data and Split

Of the 50,000 labeled samples, 40,000 (80%) are used for training, 5,000 (10%) for validation, and 5,000 (10%) for testing. Example code-mixed input: "Mala movie bhari lagi, super acting tho bro!" (Marathi–English mix). All three sentiment classes (positive, negative, neutral) are represented in each split.

b. Evaluation Metrics

The model is evaluated using accuracy, precision, recall, F1-score, and a confusion matrix:

$$Accuracy = Correct\ Predictions / Total\ Predictions \quad (5)$$

$$Precision = TP / (TP + FP), \quad Recall = TP / (TP + FN) \quad (6)$$

$$F1 = 2 \cdot Precision \cdot Recall / (Precision + Recall) \quad (7)$$

c. Experimental Procedure

Training used batch size 32, learning rate 1e-5, and the joint loss of Eq. (4), optimized with Adam over 5 epochs, with validation-set evaluation after each epoch to guide early stopping and hyperparameter selection. The final model was evaluated once on the held-out test set (5,000 samples) using the metrics in Eqs. (5)–(7) and a confusion matrix.

V. Results and Discussion

Table 2 compares the proposed hybrid model against mBERT, MuRIL, and XLM-R baselines on the held-out test set; Fig. 3 visualizes the same comparison.

Table 2. Comparison of Proposed Model with Baseline Models

Model	Accuracy	Precision	Recall	F1-Score
Proposed Hybrid Model	87.3%	86.1%	88.2%	87.1%
Baseline (mBERT)	75.4%	73.2%	74.1%	73.6%
MuRIL	80.2%	78.5%	80.0%	79.2%
XLM-R	81.5%	80.3%	81.0%	80.6%

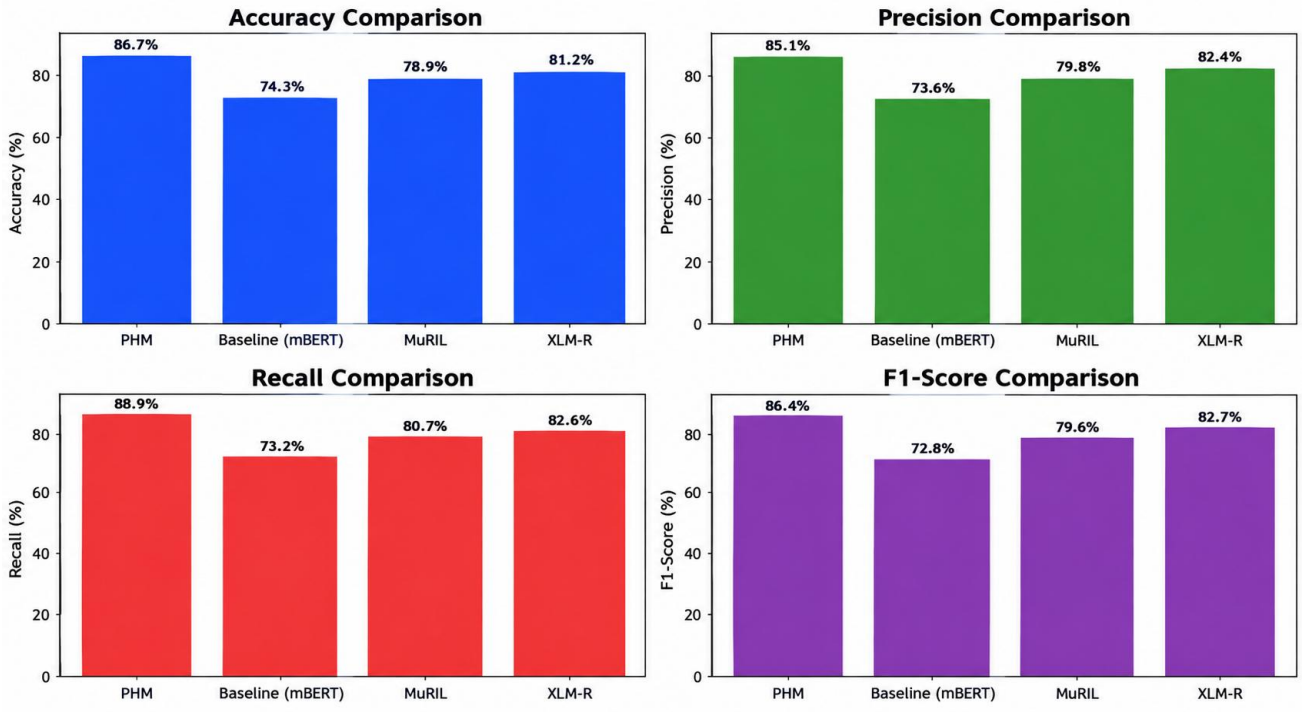

Figure 3. Accuracy, precision, recall, and F1-score: proposed model vs. mBERT, MuRIL, and XLM-R baselines.

The proposed model improves accuracy by 11.9 points over mBERT, 7.1 points over MuRIL, and 5.8 points over XLM-R, with consistent gains in precision, recall, and F1-score. MuRIL, despite being designed for Indian languages, underperforms relative to the proposed model on this tri-lingual task, consistent with its bilingual Hindi–English design focus. XLM-R's broader cross-lingual pretraining narrows the gap versus mBERT and MuRIL but still falls short of the language-specific-expert design used here.

Table 3 and Fig. 4 report the confusion matrix for the proposed model on the test set.

Table 3. Confusion Matrix for the Proposed Model

True \ Predicted	Positive	Neutral	Negative
Positive	1,310	124	56
Neutral	89	1,438	113
Negative	45	104	1,413

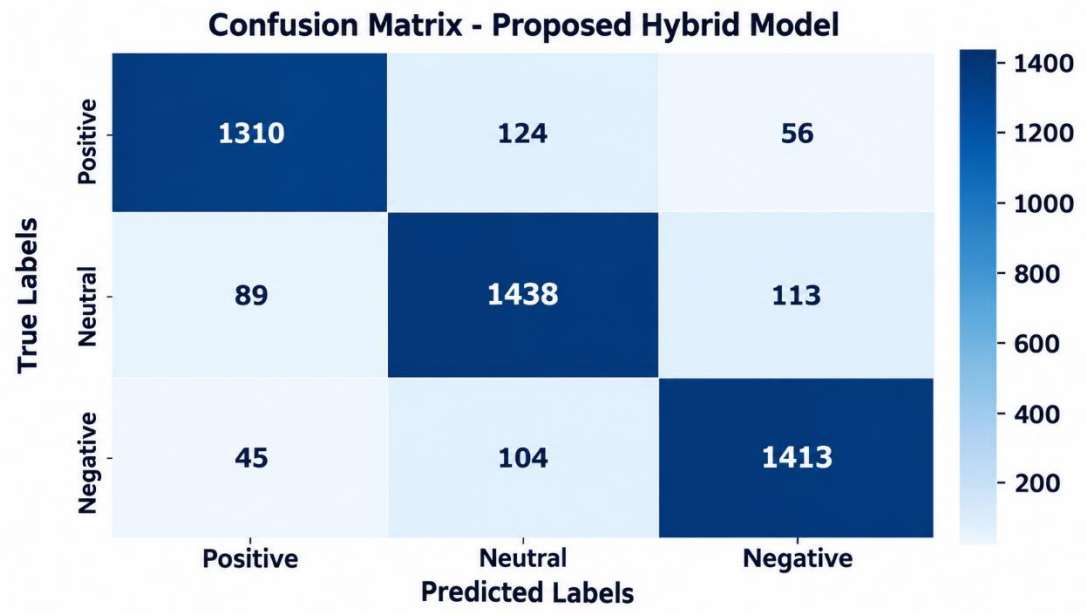

Figure 4. Confusion matrix of the proposed model on the test set.

The confusion matrix shows that most misclassifications occur between neutral and the two polar classes (124 positive→neutral, 113 neutral→negative) rather than between positive and negative directly (56 and 45 respectively), suggesting the model's main remaining difficulty is calibrating the boundary between neutral and mildly polarized sentiment — a common challenge in informal, code-mixed text where sentiment intensity is often understated.

VI. Conclusion and Future Work

This paper presented a hybrid model for tri-lingual sentiment analysis of code-mixed Marathi, Hindi, and English text, combining mBERT with an entropy-guided script-aware attention mechanism and three language-specific experts. The model outperformed mBERT, MuRIL, and XLM-R baselines in accuracy, precision, recall, and F1-score on a newly curated 50,000-sample tri-lingual dataset, and the confusion matrix confirmed reliable classification with most residual error concentrated at the neutral/polar boundary. These results support the use of language-specific experts combined with script-aware attention for real-world multilingual applications such as social media monitoring, multilingual customer feedback analysis, and sentiment-based decision-making.

Future Work

- Sarcasm and irony detection: the current model does not identify sarcasm, common in informal code-mixed communication; a dedicated detection module using sentiment lexicons or pre-trained classifiers is a natural extension.
- Real-time deployment: the current architecture is not optimized for real-time inference; latency profiling and model compression would be needed for live social-media monitoring use cases.
- Cross-domain generalization: the model is trained on social media and customer-feedback text; sentiment expression differs in domains such as health, politics, and finance, and would require domain-specific fine-tuning or transfer learning to generalize.

REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.

[2] T. Pires, E. Schlinger, and D. Garrette, "How multilingual is multilingual BERT?," Proc. of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 4996–5001, 2019.

[3] S. Khanuja, D. Bansal, S. Mehtani, S. Khosla, A. Dey, B. Gopalan, D. K. Margam, P. Aggarwal, R. T. Nagipogu, S. Dave, S. Gupta, S. C. B. Gali, V. Subramanian, and P. Talukdar, "MuRIL: Multilingual representations for Indian languages," arXiv:2103.10730, 2021.

[4] D. Kakwani, A. Kunchukuttan, S. Golla, G. N.C., A. Bhattacharyya, M. M. Khapra, and P. Kumar, "IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages," Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 4948–4961, 2020.

[5] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," Proc. of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 8440–8451, 2020.

[6] A. Joshi, A. Prabhu, M. Shrivastava, and V. Varma, "Towards sub-word level compositions for sentiment analysis of Hindi-English code mixed text," Proc. of COLING 2016, the 26th Int. Conf. on Computational Linguistics: Technical Papers, pp. 2482–2491, 2016.

[7] P. Patwa, G. Aguilar, S. Kar, S. Pandey, S. PYKL, B. Gambäck, T. Chakraborty, T. Solorio, and A. Das, "SemEval-2020 Task 9: Overview of sentiment analysis of code-mixed tweets," Proc. of the 14th Int. Workshop on Semantic Evaluation (SemEval-2020), pp. 774–790, 2020.

[8] S. Sitaram, K. R. Chandu, S. K. Rallabandi, and A. W. Black, "A survey of code-switched speech and language processing," arXiv:1904.00784, 2019.

[9] S. Khanuja, S. Dandapat, A. Srinivasan, S. Sitaram, and M. Choudhury, "GLUECoS: An evaluation benchmark for code-switched NLP," Proc. of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 3575–3585, 2020.

[10] A. Kulkarni, M. Mandhane, M. Likhitkar, G. Kshirsagar, and R. Joshi, "L3CubeMahaSent: A Marathi tweet-based sentiment analysis dataset," Proc. of the 11th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA), pp. 213–220, 2021.

[11] N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. Le, G. Hinton, and J. Dean, "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer," arXiv:1701.06538, 2017.

[12] P. Prettenhofer and B. Stein, "Cross-language text classification using structural correspondence learning," Proc. of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 1118–1127, 2010.

[13] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," Proc. of the 3rd Int. Conf. on Learning Representations (ICLR), 2015.

[14] R. Caruana, "Multitask learning," Machine Learning, vol. 28, no. 1, pp. 41–75, 1997.

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Proc. of the 3rd Int. Conf. on Learning Representations (ICLR), 2015.

[16] A. Shelotkar and S. K. Oruganti, "Enhancing multilingual sentiment analysis with large language models: Current trends and future direction," Proc. Vol. 1 No. 2 (2026), LGPR, Malaysia.

[17] A. Shelotkar and S. K. Oruganti, "Entropy-guided script-aware mixture-of-experts to tri-lingual sentiment analysis of Marathi–Hindi–English communication," Proc. Vol. 1 No. 4 (2026), LGPR, Malaysia.