Domain-Adaptive Multitask BERT with Graph Context Modeling for Code-Mixed Hinglish Sentiment Classification

Rahul Mishra
Mechanical Engineering Department, Rungta International Skills University, Bhilai Chhattisgarh, 490024, India | Mechanical Engineering Department, Rungta College of Engineering and Technology, Bhilai Chhattisgarh, 490024, India
mishrarahul87@gmail.com

A. Kaliappan
School of Computing, SRM Institute of Science and Technology, Tiruchirappalli, Tamil Nadu, India
kaliappantpr@gmail.com

Sarala D. V.
Computer Science and Engineering Department, BMS College of Engineering, Bangalore, India
saraladv.cse@bmsce.ac.in

T. V. Hyma Lakshmi
Department of Electronics and Communication Engineering, Sagi Rama Krishnam Raju Engineering College, Bhimavaram, Andhra Pradesh, 534204, India
tvhymalakshmi@srkrec.edu.in

Prabhat Kumar Ravi
University Institute of Computing, Chandigarh University, Mohali, Punjab, India
dr.prabhatravi@gmail.com

Y. Harika Devi
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, Telangana, India
harikayelaga11@gmail.com

Satyajee Srivastava
Department of Computer Science and Engineering, School of Engineering and Technology, Manav Rachna International Institute of Research and Studies (Deemed to be University), Faridabad, Haryana, India
drsatyajee@gmail.com

Deepak Suresh Asudani
Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India
deepak.s.asudani@gmail.com

Sanjana M. Nagaraj
Department of Computer Science and Business Systems, Dayananda Sagar College of Engineering, Bengaluru, India
sanjanamnagaraj@gmail.com

Yogesh H. Bhosale
Department of Computer Science and Engineering, CSMSS CHH. Shahu College of Engineering, Chhatrapati Sambhajinagar, India
yogeshbhosale988@gmail.com (corresponding author)

Received: 29 November 2025 | Revised: 27 December 2025 | Accepted: 3 January 2026

 DOI: https://doi.org/10.48084/etasr.16589

ABSTRACT

Hinglish, a widely used Hindi–English code-mixed language on social media, presents unique challenges for sentiment analysis due to transliteration variability, script mixing, and inconsistent grammar. To address these issues, this study proposes the Cross-Lingual Domain-Adaptive Multitask Graph-Enhanced Bidirectional Encoder Representations from Transformers (CDMG-BERT), tailored for code-mixed sentiment classification. The model integrates four key components: English-to-Hinglish embedding alignment, domain-adversarial training using 92k English samples, multitask learning with auxiliary POS tagging, and graph-based token relational modeling for long-range contextual refinement. Experimental results on a 12,000-sample IIT Bombay Hinglish dataset show that CDMG-BERT outperforms mBERT, XLM-R, and deep learning baselines, with an F1-score of 84.8%. Ablation analysis suggests that each architectural module works well, with cross-lingual alignment and domain adaptation providing the largest gains. These results demonstrate that the model is robust enough to handle spelling differences, code-mixing intensity, and inconsistencies between the Roman and Devanagari scripts. This makes CDMG-BERT a good choice for sentiment analysis in a multilingual environment with limited resources.

Keywords-Hinglish; sentiment analysis; code-mixing; cross-lingual learning; domain adaptation; transformer models

I. INTRODUCTION

Hinglish is a Hindi–English code-mixed language widely used on social media in India and among Indian diaspora communities, where lexical items from both languages appear within the same sentence or discourse. It is predominantly written in Roman (Latin) script, although occasional Devanagari characters may appear, with Romanization being preferred due to keyboard accessibility and informal online practices. Romanized Hindi lacks standard spelling conventions, resulting in multiple variants for the same word (e.g., “accha,” “acha,” “achha” for अच्छा). Hinglish text also exhibits script mixing, spelling variability, informal grammar, emojis, elongated words, and symbolic expressions, alongside diverse code-mixing patterns such as intra-sentential mixing and lexical borrowing. These characteristics introduce significant linguistic noise, making Hinglish sentiment analysis substantially more challenging than monolingual text processing. With the increasing use of social media in India, Hinglish content has grown substantially.

Hindi and English are mixed together in the same sentence and mostly written in Roman script [1]. Multilingual transformers, like mBERT and XLM-R, have improved cross-lingual semantic modeling, but they still perform poorly on Hinglish, as they have not been specifically trained on code-mixed patterns or Romanized Hindi expressions [2]. This difference stresses the need for a special model that can handle the unique structure and dynamics of Hinglish text. Current methods, especially conventional machine learning models, such as SVMs and XGBoost, predominantly depend on superficial lexical features, which are inadequate for elucidating profound semantic and syntactic relationships in code-mixed text. Neural architectures, like BiLSTMs and CNN–LSTM hybrids, perform better in representing context.

Authors in [4] introduced the L3Cube-HingCorpus and the HingBERT model, developed specifically for Hindi–English code-mixed data. Their study highlighted major Hinglish challenges, including heavy code-mixing, transliteration inconsistencies, and the absence of dedicated pretraining resources. HingBERT significantly outperformed multilingual baselines, achieving notable improvements of 4-7% in F1-score compared to standard mBERT and XLM-R on Hinglish benchmarks. Authors in [5] focused on adapting mBERT for English–Hindi code-mixed sentiment analysis. They investigated issues such as domain mismatch, ambiguous sentiment cues, and noise in Romanized Hindi tokens. Their approach relied on mBERT finetuning with optimized tokenization and contextual embedding extraction. The results showed that mBERT achieved an F1-score of approximately 78–80%, demonstrating strong cross-lingual transfer capabilities. Authors in [6] explored Graph Neural Networks (GNNs) for text classification. Their findings indicated that GNN-based architectures yielded performance gains of 3–6% on datasets containing hierarchical or long-range dependencies, making them highly suitable for complex code-mixed sentences. Authors in [7] proposed an adversarial training strategy for robust text classification. Their method employed a Gradient Reversal Layer (GRL) and adversarial perturbations to learn domain-invariant representations. Experimental evaluations showed that adversarially trained models outperformed baseline CNN/BiLSTM classifiers by 2–5% in accuracy, particularly under noisy or cross-domain shifts. Authors in [8] investigated pre-trained transformer models for multilingual code-mixed languages involving Indonesian, Javanese, and English. The proposed model achieved consistent improvements of 5–8% in F1-score over multilingual baselines.

To address these issues, the present study creates a resilient transformer-based framework specifically designed for Hinglish sentiment analysis [3]. The study aims to conduct a detailed assessment of the proposed model against classical, deep learning, and multilingual benchmarks using a real Hinglish sentiment dataset to confirm its performance enhancements. The main contribution of this research lies in the creation of CDMG-BERT, a Cross-lingual Domain-adaptive Multitask Graph-enhanced Transformer engineered for code-mixed sentiment classification. The model presents a projection-based embedding alignment mechanism that translates English representations into the Hinglish embedding space, effectively resolving transliteration discrepancies.

II. METHODOLOGY

This study follows a structured end-to-end pipeline that transforms raw Hinglish text into sentiment predictions through a sequence of well-defined stages. First, the raw code-mixed dataset is pre-processed and tokenized using a Sentence Piece tokenizer to handle transliteration variability. Next, cross-lingual embedding alignment maps English representations into a Hinglish embedding space to mitigate vocabulary mismatch. Domain-adversarial training is then employed using large-scale English sentiment corpora to learn domain-invariant features. The aligned representations were jointly optimized through a multitask learning framework that performs sentence-level sentiment classification and token-level POS tagging. Figure 1 illustrates the workflow of the proposed CDMG-BERT for sentiment analysis on the IIT Bombay Hinglish Sentiment Dataset.

A. Overview of the CDMG-BERT Framework

Given a Hinglish input sequence $X = \{x_1, x_2, \dots, x_n\}$, the objective of the current work is to jointly predict (a) sentence-level sentiment $y_s \epsilon \{Positive, \text{Negative}, Neutral\}$ and (b) token-level POS tags $Y_p = \{p_1, p_2, \dots, p_n\}$.

B. Cross-Lingual Representation Alignment

The process begins with a pre-trained English BERT model $BERT_{EN}$. Since Hinglish includes romanized Hindi words not present in the English vocabulary, a projection-based embedding alignment [9] is learned, where a trainable linear projection $W \in \mathbb{R}^{d\times d}$ maps English embeddings into Hinglish space as given in:

$$E_{HI}(w) = W \cdot E_{EN}(w) \tag{1}$$

where $E_{EN}(w)$is the English BERT embedding for the token w and $E_{HI}(w)$ is the Hinglish embedding initialized randomly. The alignment loss is computed over a pivot dictionary D_{pivot} of English–Hinglish transliteration pairs [10] as defined in:

$$L_{align} = \sum_{(w_{en}, w_{hi}) \in D_{pivot}} \| WE_{EN}(w_{en}) - E_{HI}(w_{hi}) \|^2 \tag{2}$$

This ensures semantic consistency between English and Hinglish vocabularies.

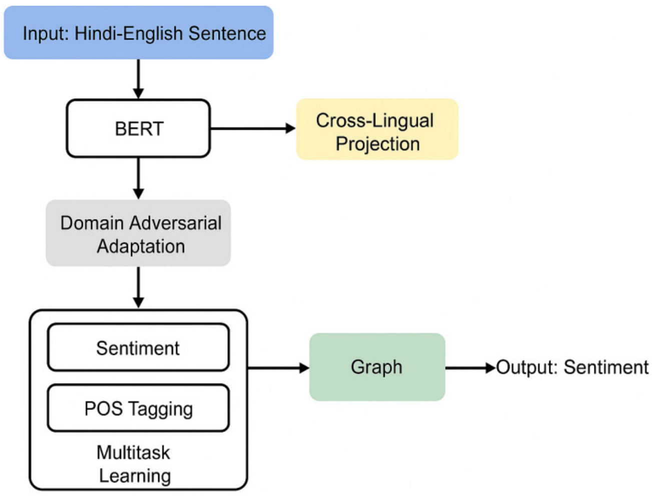

Fig. 1. CDMG-BERT for sentiment analysis.

C. Domain-Adversarial Adaptation

The domain-adversarial adaptation minimizes the distribution gap between source and target domains by learning domain-invariant features [11]. The adversarial domain loss is defined in:

$$L_{domain} = -\sum_{(X,d)}[dlog\ D(F(X)) + (1-d)log\ (1 - D(F(X)))] \tag{3}$$

where D_s is the source domain (English datasets such as IMDb, SST), D_t is the target domain (IIT Bombay Hinglish data), $F(X)$ represents a shared BERT encoder for Domain-Adversarial Neural Network (DANN), C_s is the sentiment classifier, D is the domain discriminator, and $d\ \epsilon \{0 =$ English, 1= Hinglish}represents the domain label.

1) GRL

GRL multiplies the gradient by –λ, forcing the encoder to learn domain-invariant features [12], as defined in:

$$\theta_F \leftarrow \theta_F - \lambda \frac{\partial L_{domain}}{\partial \theta_F} \tag{4}$$

D. Multitask Learning: Sentiment and POS Tagging

A multitask learning framework that jointly trains the model for sentiment classification and POS tagging was employed, enabling shared representations that improve generalization across tasks. This design follows established multitask NLP principles and advances in multitask sentiment modeling [13].

1) Sentiment Classification

Given the sentence representation $h = F(X)$, as defined in (5), the predicted sentiment label is computed as:

$$\hat{y}_s = \text{softmax}(W_s h + b_s) \tag{5}$$

The sentiment loss is calculated using (6).

$$L_{sent} = -\sum_{c=1}^{C} y_s(c) log\ \hat{y}_s(c) \tag{6}$$

2) POS Tag Prediction

For each token $x_i$with representation h_i, as defined in (7), the predicted POS tag distribution is computed as:

$$\hat{p}_i = \text{softmax}(W_p h_i + b_p) \quad (7)$$

The POS tag is calculated by:

$$L_{pos} = -\sum_{i=1}^{n} \sum_{k=1}^{K} y_{p_i}(k) log\ \hat{p}_i(k) \quad (8)$$

3) Joint Multitask Loss

The joint multitask learning loss is defined as:

$$L_{MTL} = \alpha L_{sent} + \beta L_{pos} \quad (9)$$

where α and β control task importance [14].

E. Graph-Based Context Modeling for Code-Mixed Sentences

Hinglish sentences often contain long-range and cross-lingual dependencies [15]. To capture these, a token graph was constructed:

$$G = (V, E)$$

where V represents the tokens, and E denotes the edges based on semantic similarity or co-occurrence. Edge weights are based on embedding similarity as defined in:

$$w_{ij} = exp\,(-\| E_{HI}(x_i) - E_{HI}(x_j) \|^2) \quad (10)$$

1) GNN Encoding

A Graph Convolutional Network (GCN) was applied as shown in:

$$H^{(l+1)} = \sigma(\widetilde{D}^{-1/2} \tilde{A} \widetilde{D}^{-1/2} H^{(l)} W^{(l)} \quad (11)$$

where $\tilde{A} = A + I$ represents the adjacency with self-loops, while $H^{(0)} = F(X)$ represents BERT token embeddings. The graph-enhanced sentence representation is shown in:

$$h_{GNN} = meanpool(H^{(L)}) \quad (12)$$

F. Fusion Layer and Final Prediction

The local BERT and global GNN relational features were combined as shown in:

$$h_{fusion} = \gamma h_{BERT} + (1 - \gamma) h_{GNN} \quad (13)$$

The final sentiment prediction was conducted:

$$\hat{y} = softmax(W_f h_{fusion} + b_f) \quad (14)$$

G. Total Objective Function

The overall training objective integrates cross-lingual alignment, domain alignment, multitask learning, and graph-enhanced sentiment prediction, as defined by the final Prediction Function.

$$L_{total} = \lambda_1 L_{align} + \lambda_2 L_{domain} + \lambda_3 L_{MTL} + \lambda_4 L_{sent}(fusion)$$

III. EXPERIMENTAL SETTINGS

A. Data Preprocessing

The Hinglish text underwent a structured preprocessing pipeline to ensure high-quality model input. The primary dataset used in this study consists of 12,000 code-mixed Hinglish reviews from the IIT Bombay Hinglish Sentiment Corpus, a publicly available benchmark dataset designed for Hindi English sentiment analysis [16]. To support domain-adversarial training, an additional ≈92,000 English sentiment samples from the IMDb and SST-2 datasets [19] were incorporated as the source domain. For the extra POS-tagging task, all Hinglish samples got POS tags [17].

B. Experimental Setup

A standard experimental setup was used to train the model. The dataset was randomly divided into three mutually exclusive subsets: 70% for training, 10% for validation, and 20% for testing, while preserving class distribution to avoid evaluation bias. The AdamW optimizer was used for training, with a learning rate of $3{\times}10^{-5}$, a batch size of 32, and a total of 50 epochs. The main optimization goal was to minimize cross-entropy loss.

C. Baseline Models

Multiple competitive baseline models, including the BiLSTM model, which uses word embeddings and bidirectional recurrent structures, as well as a hybrid CNN–LSTM, which combines local semantic feature extraction [18], were employed to maintain the efficacy and novelty of the proposed CDMG-BERT framework.

IV. RESULTS AND DISCUSSION

A. Performance Comparison

Table I presents the performance of different models on the Hinglish sentiment dataset. The results show that classical models, such as XGBoost, had moderate performance, with an F1-score of 66.3%, as they are unable to classify semantic and contextual cues in code-mixed Hinglish. Other classical models improved performance with an F1-score of 71–73%, especially XLM-R with an F1 score of 80.1%. Table II summarizes the class-wise performance of the CDMG-BERT model based on accuracy, precision, recall, and F1-score. Figures 2 and 3 depict the training and validation curves for accuracy and loss of CDMG-BERT, respectively. Both curves exhibit a steady upward trend, indicating stable learning and the generalization ability without overfitting. Figure 3 shows how the training and validation loss changed over 50 epochs.

TABLE I. PERFORMANCE OF MODELS ON HINGLISH SENTIMENT DATASET

Model	Accuracy	Precision	Recall	F1-score
XGBoost (TF-IDF)	68.4%	67.1%	65.9%	66.3%
BiLSTM	72.8%	71.9%	70.2%	71.0%
CNN–LSTM Hybrid	74.1%	73.5%	72.8%	73.1%
mBERT(fine-tuned)	78.6%	78.1%	77.4%	77.8%
XLM-R(fine-tuned)	80.9%	80.5%	79.8%	80.1%
CDMG-BERT (Proposed)	85.7%	85.1%	84.6%	84.8%

TABLE II. CLASS-WISE PERFORMANCE OF CDMG-BERT

Sentiment class	Precision	Recall	F1-score
Positive	87.3%	86.2%	86.7%
Negative	84.9%	84.1%	84.5%
Neutral	83.1%	82.6%	82.8%

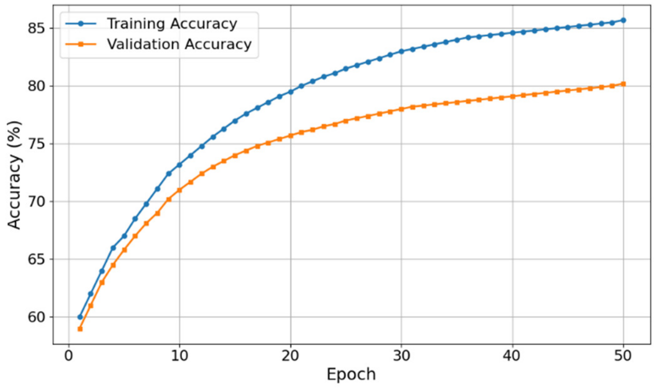

Fig. 2. Training and validation accuracy for CDMG-BERT.

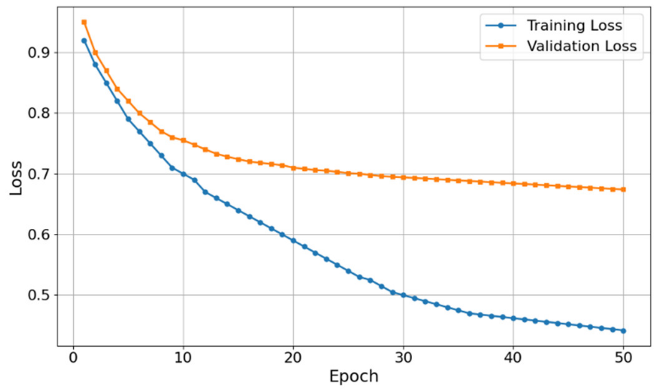

Fig. 3. Training and validation loss for CDMG-BERT.

Table III displays a summary of the results by class. CDMG-BERT achieves the highest performance on the Positive class with an F1-score of 86.7%, followed by the Negative class (84.5%) and the Neutral class (82.8%). As expected, the model has poor performance on neutral sentences because neutral expressions in Hinglish often have subtle or unclear cues ("Theek hai," "Thoda sahi tha," "Okay-ish"), which makes it harder to differentiate between weakly positive or weakly negative statements. Figure 4 illustrates a stacked confusion comparison of mBERT, XLM-R, and CDMG-BERT.

TABLE III. ABLATION STUDY OF CDMG-BERT COMPONENTS

Configuration	Accuracy	F1-score
Full model (CDMG-BERT)	85.7%	84.8%
Without the GNN graph layer	83.2%	82.5%
Without POS, multitask learning	82.8%	82.1%
Without domain adversarial training	80.6%	79.9%
Without cross-lingual alignment	78.4%	77.6%

B. Ablation Study and Component Contribution

Table IV presents the results of the ablation study, illustrating the contribution of each architectural component to overall performance. Removing the GNN module reduces the F1-score by more than 2.3%, which shows that graph-based relational modeling is important for capturing long-range dependencies and code-mixed token interactions. Table IV further examines the effect of the GNN layer by assessing performance across three different types of Hinglish sentences. When GNN layers are added, sentences with long dependencies and a large code mixing show the greatest improvements, with gains of up to 8.7%.

TABLE IV. IMPACT OF GNN LAYERS ON DIFFERENT SENTENCE TYPES

Sentence Type	Without GNN	With GNN	Improvement
Hinglish with long dependencies	74.1%	82.8%	+8.7%
Romanized Hindi dominant sentences	70.5%	79.2%	+8.7%
English-heavy sentences	80.6%	83.4%	+2.8%

As portrayed in Figure 4, CDMG-BERT consistently outperforms baseline models across all evaluation metrics, demonstrating improved robustness on noisy code-mixed Hinglish text. While classical and deep learning models struggle with surface features and high code-mixing intensity, the proposed architecture leverages cross-lingual alignment, domain adaptation, and graph-based modeling to achieve more stable and reliable sentiment inference.

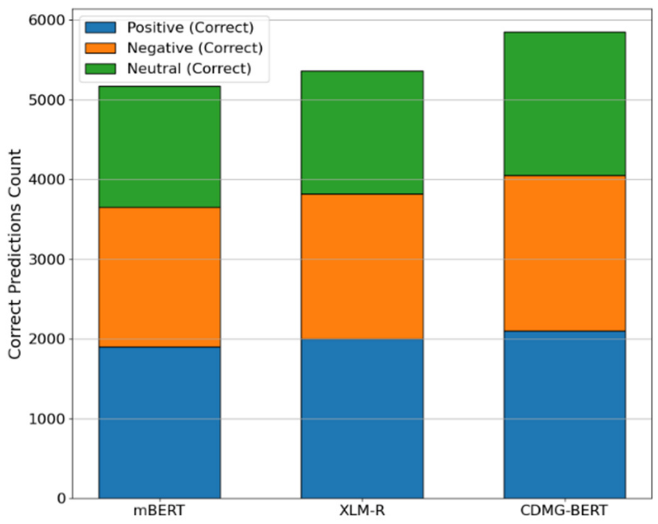

Fig. 4. Stacked confusion comparison.

V. CONCLUSION

The current study presented a Cross-Lingual Domain-Adaptive Multitask Graph-Enhanced Transformer (CDMG-BERT), augmented with graph-based Graph Neural Network (GNN) contextual modeling for Hinglish sentiment analysis. The proposed framework effectively addresses key challenges inherent to code-mixed Hinglish text, including transliteration variability, script inconsistency, domain mismatch, and long-range token dependencies. By integrating cross-lingual embedding alignment, domain-adversarial training, auxiliary POS-tagging supervision, and graph-based relational modeling, CDMG-BERT learns robust and linguistically informed representations for sentiment classification. Extensive experiments on the IIT Bombay Hinglish Sentiment Dataset show that CDMG-BERT consistently outperforms classical, deep learning, and multilingual transformer baselines, achieving an F1-score of 84.8%. Ablation studies and confusion analysis confirm the contribution of each architectural component and demonstrate robust generalization across varying code-mixing intensities and noisy social media text. Future work will explore large-scale Hinglish pretraining, sarcasm-aware sentiment modeling, and multimodal extensions incorporating emojis and visual cues.

REFERENCES

[1] L. Nguyen, O. Mayeux, and Z. Yuan, "Code-Switching Input for Machine Translation: A Case Study of Vietnamese–English Data," *International Journal of Multilingualism*, vol. 21, no. 4, pp. 2268–2289, Oct. 2024, https://doi.org/10.1080/14790718.2023.2224013.

[2] M. S. U. Miah, M. M. Kabir, T. B. Sarwar, M. Safran, S. Alfarhood, and M. F. Mridha, "A Multimodal Approach to Cross-Lingual Sentiment Analysis with Ensemble of Transformer and LLM," *Scientific Reports*, vol. 14, no. 1, Apr. 2024, Art. no. 9603, https://doi.org/10.1038/s41598-024-60210-7.

[3] Z. Cao, Y. Zhou, A. Yang, and S. Peng, "Deep Transfer Learning Mechanism for Fine-Grained Cross-Domain Sentiment Classification," *Connection Science*, vol. 33, no. 4, pp. 911–928, Oct. 2021, https://doi.org/10.1080/09540091.2021.1912711.

[4] R. Nayak and R. Joshi, "L3Cube-HingCorpus and HingBERT: A Code Mixed Hindi-English Dataset and BERT Language Models." arXiv, 2022, https://doi.org/10.48550/ARXIV.2204.08398.

[5] S. K. Singh, A. Sharma, Sahil, D. Singh, S. Pandit, and U. Saghir, "Sentiment Analysis of English-Hindi Code-Mixed Text Using mBERT Model," in *2025 3rd International Conference on Inventive Computing and Informatics*, Bangalore, India, June 2025, pp. 552–556, https://doi.org/10.1109/ICICI65870.2025.11069692.

[6] K. Wang, Y. Ding, and S. C. Han, "Graph Neural Networks for Text Classification: a Survey," *Artificial Intelligence Review*, vol. 57, no. 8, July 2024, Art. no. 190, https://doi.org/10.1007/s10462-024-10808-0.

[7] X. Liu, S. Dai, G. Fiumara, and P. De Meo, "An Adversarial Training Method for Text Classification," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 8, Sept. 2023, Art. no. 101697, https://doi.org/10.1016/j.jksuci.2023.101697.

[8] A. F. Hidayatullah, R. A. Apong, D. T. C. Lai, and A. Qazi, "Pre-Trained Language Model for Code-Mixed Text in Indonesian, Javanese, and English Using Transformer," *Social Network Analysis and Mining*, vol. 15, no. 1, Mar. 2025, Art. no. 30, https://doi.org/10.1007/s13278-025-01444-9.

[9] Z. Chi *et al.*, "InfoXLM: An Information-Theoretic Framework for Cross-Lingual Language Model Pre-Training," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Online, 2021, pp. 3576–3588, https://doi.org/10.18653/v1/2021.naacl-main.280.

[10] M. Ulčar and M. Robnik-Šikonja, "Cross-Lingual Alignments of ELMo Contextual Embeddings," *Neural Computing and Applications*, vol. 34, no. 15, pp. 13043–13061, Aug. 2022, https://doi.org/10.1007/s00521-022-07164-x.

[11] F. Hou *et al.*, "Gradient-Aware Domain-Invariant Learning for Domain Generalization," *Multimedia Systems*, vol. 31, no. 1, Feb. 2025, Art. no. 40, https://doi.org/10.1007/s00530-024-01613-4.

[12] L. Zhang, X. Wei, F. Yang, C. Zhao, B. Wen, and Y. Lu, "Cross-Domain Sentiment Classification with Mere Contrastive Learning and Improved Method," in *2024 3rd International Conference on Artificial Intelligence and Computer Information Technology (AICIT)*, Yichang, China, Sept. 2024, pp. 1–10, https://doi.org/10.1109/AICIT62434.2024.10730527.

[13] N. Sharma and B. Verma, "Recent Advances in Transfer Learning for Natural Language Processing (NLP)," in *A Handbook of Computational Linguistics: Artificial Intelligence in Natural Language Processing*, Y. B. Singh, A. D. Mishra, P. Singh, and D. K. Yadav, Eds. Bentham Science Publishers, 2024, pp. 228–254.

[14] S. Zhang, C. Yin, and Z. Yin, "Multimodal Sentiment Recognition with Multi-Task Learning," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 1, pp. 200–209, Feb. 2023, https://doi.org/10.1109/TETCI.2022.3224929.

[15] Z. Ye, Y. J. Kumar, G. O. Sing, F. Song, and J. Wang, "A Comprehensive Survey of Graph Neural Networks for Knowledge Graphs," *IEEE Access*, vol. 10, pp. 75729–75741, 2022, https://doi.org/10.1109/ACCESS.2022.3191784.

[16] A. Kunchukuttan, P. Mehta, and P. Bhattacharyya, "The IIT Bombay English–Hindi Parallel Corpus," in *Eleventh International Conference on Language Resources and Evaluation*, Miyazaki, Japan, May 2018.

[17] S. Choo and W. Kim, "A Study on the Evaluation of Tokenizer Performance in Natural Language Processing," *Applied Artificial Intelligence*, vol. 37, no. 1, Dec. 2023, Art. no. 2175112, https://doi.org/10.1080/08839514.2023.2175112.

[18] S. Perumal and K. Kathirvelu, "Enhancing the Quality of Service in Video Game Live Streaming Using Big Data Analytics with DNN Classification and BERT-Based Sentiment Analysis," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25426–25431, Aug. 2025, https://doi.org/10.48084/etasr.11495.

[19] R. Socher *et al.*, "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," in *Conference on Empirical Methods in Natural Language Processing*, Seattle, WA, USA, Oct. 2013, pp. 1631–1642.