Indiana Journal of Multidisciplinary Research
Abbreviated Tittle- Ind J Multi Res
ISSN *(Online)*- 2583-3820
Journal Homepage Link- https://indianapublications.com/Journals/IJMR
DOI: https://doi.org/10.5281/zenodo.21601936

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Article

Volume-06|Issue-04|2026

Hinglish-to-English Translation Using NLP and Machine Learning Techniques

Aditya Singh[1], Aiman Sabah[2], Apeksha Jawali[3], Gyanada Muvvala[4], Prof. Samatha R. Swamy[5]

[1,2,3,4]Student, Department of Information Science and Engineering, RV Institute of Technology and Management (RVITM), Bengaluru, India,
[5]Assistant Professor, Department of Information Science and Engineering, RV Institute of Technology and Management (RVITM), Bengaluru, India

Article History
Received: 05.05.2026
Accepted: 22.06.2026
Published: 25.07.2026

Citation
Singh, A., Sabah, A., Jawali, A., Muvvala, G., Swamy, S. R. (2026). Hinglish-to-English Translation Using NLP and Machine Learning Techniques. *Indiana Journal of Multidisciplinary Research, 6*(4), 192-197.

Abstract: In a multilingual society like India, communication often involves code-mixing, particularly Hinglish, a blend of Hindi and English. This poses challenges for traditional Natural Language Processing (NLP) systems due to transliteration, phonetic variations, and informal grammar. This study proposes an NLP-based Hinglish-to-English translation system leveraging datasets such as PHINC and CMU Hinglish Dog. The system integrates language identification, transliteration, dependency parsing, and neural machine translation using models such as mBART, mT5, and IndicTrans. Three approaches—namely, a multistage pipeline, hybrid models, and reinforcement learning-based translation—are explored. The results indicate improved translation accuracy and contextual understanding. The proposed system demonstrates potential applications in conversational AI, education, healthcare, and content moderation.

Keywords: Hinglish Translation, Code-Mixed Language, NLP, Neural Machine Translation, Transformer Models

INTRODUCTION

In today's digitally connected and linguistically diverse world, communication often transcends traditional language boundaries. In countries such as India, users frequently switch between languages within a single sentence, a phenomenon commonly known as **code-mixing** [1]. One prominent example is **Hinglish**, a blend of Hindi and English that has become a widely used medium of informal communication across social media platforms, messaging applications, and conversational interfaces [10]. Despite its widespread usage, Hinglish presents significant challenges to Natural Language Processing (NLP) systems because of its unstructured nature, phonetic spellings, transliterations, and inconsistent grammar [6], [14].

Existing machine translation models often struggle to interpret and translate code-mixed text accurately, resulting in poor comprehension and reduced effectiveness of language-based AI applications [7]. This work addresses this limitation by developing a Hinglish-to-English translation system using advanced NLP and machine learning techniques. The primary objective is to build a robust, context-aware translation model capable of handling transliterations, identifying language segments, and converting code-mixed sentences into grammatically correct English.

By leveraging large annotated datasets such as **PHINC** and **CMU Hinglish Dog**, and employing state-of-the-art transformer-based models including **mBART** and **IndicTrans** [16], [17], together with hybrid translation pipelines, the proposed system aims to deliver accurate, scalable, and contextually appropriate translations. The system has potential applications in conversational AI, content moderation, education, healthcare, and government services. Future enhancements, including speech-to-text integration and multilingual support, can further bridge linguistic barriers and promote digital inclusivity for users of code-mixed languages.

LITERATURE SURVEY

The increasing prevalence of code-mixed languages, particularly Hinglish, across digital platforms has attracted considerable research interest in developing effective machine translation (MT) systems. One foundational study introduced a gold-standard parallel corpus comprising **6,096 English-Hindi code-mixed sentences** translated into English by bilingual experts. The work also presented an augmentation pipeline compatible with standard MT systems such as **GNMT** and **Bing Translator**, resulting in notable improvements in BLEU and WER metrics [6]. The proposed approach emphasized the importance of tailored preprocessing strategies and laid the foundation for future end-to-end hybrid translation systems integrating normalization, language detection, and translation into a unified workflow.

***Corresponding Author:** Prof. Samatha R. Swamy

In another domain, **Schäfer *et al.*** proposed a cross-lingual annotation transfer framework for Named Entity Recognition (NER) in German clinical texts using English-trained models [2]. Their dual-mode pipeline, incorporating neural machine translation and multilingual contextual embeddings, demonstrated that backward translation (German to English) followed by annotation realignment produced superior performance, particularly in medication recognition tasks. This study highlighted the adaptability of English-trained models to low-resource languages through careful alignment strategies, although challenges related to annotation inconsistencies and linguistic divergence remain.

Addressing the absence of parallel corpora in e-commerce applications, **Kulkarni and Garera** [3] employed an unsupervised domain adaptation strategy using **mBART-50** to translate Hindi search queries into English. Their framework combined denoising autoencoding, cross-lingual training (**CrossLT**), and adversarial embedding alignment. CrossLT improved BLEU scores by more than **20 points**, with additional gains achieved through minimal supervised fine-tuning. This scalable, domain-adaptive framework demonstrates promising potential for low-resource machine translation in practical search applications.

To address script variation in multilingual pretrained models, the **TRANSLICO** framework [5] introduced contrastive learning combined with transliteration to align representation spaces across scripts such as **Devanagari, Arabic, and Latin**. Through joint optimization using masked language modeling and contrastive learning objectives, the fine-tuned **FURINA** model outperformed baseline approaches in sentence retrieval and sequence labeling tasks, particularly for Indic languages. Unlike previous romanization-based approaches, TRANSLICO preserved script fidelity, making it a robust, script-independent solution for cross-script language generalization.

The **GUI** system presented at **MixMT 2022** approached Hinglish machine translation bidirectionally by translating between English/Hindi and Hinglish. Using **mBART** and **OPUS-MT**, the framework incorporated vocabulary trimming and transliteration preprocessing to improve translation quality. Additional post-processing techniques resulted in significant improvements in ROUGE-L and WER scores [25]. Although effective, the framework relied heavily on synthetic datasets and encountered difficulties with highly informal social media text, highlighting the need for richer real-world datasets and improved handling of informal linguistic structures.

Building upon pretrained transformer architectures, a recent study applied **mBART** and **mT5** [16] within a dual curriculum learning framework to improve Hinglish-to-English translation. The models were initially trained on English-to-Hinglish translation tasks, enabling them to develop familiarity with code-mixed sentence structures before fine-tuning for Hinglish-to-English translation. The **mT5** model achieved a **BLEU score of 29.5**, nearly doubling the baseline performance and demonstrating the effectiveness of staged training. Although limited by dataset diversity, this approach provides a strong foundation for future improvements through data augmentation and hybrid rule-based and neural translation methods.

Another study examined both the linguistic and technical challenges associated with processing Hinglish, a widely used combination of Hindi and English on Indian social media platforms. Due to its informal writing style, inconsistent grammar, and frequent language switching, Hinglish presents considerable challenges for conventional machine translation systems. To overcome these issues, **Wu *et al.*** proposed a customized translation and classification pipeline specifically designed for Hinglish [14]. Their framework emphasizes key preprocessing tasks such as transliteration of Romanized Hindi, text normalization, and context-aware ambiguity resolution. By integrating accurate language identification with bilingual dictionary support, the proposed approach outperformed general-purpose systems such as **Google Translate**, as demonstrated through BLEU score evaluations. This work represents an important step toward developing adaptive and accurate mixed-language processing systems through the integration of rule-based methods and deep learning techniques.

The literature also highlights the **mT5** model, an extension of the original **T5** architecture specifically designed for multilingual applications. Trained on the multilingual **Common Crawl (mC4)** dataset [16], which spans **101 languages**, mT5 adopts a unified text-to-text learning framework capable of handling a broad range of natural language processing tasks. Its balanced multilingual training strategy ensures competitive performance across both high-resource and low-resource languages. The model has achieved state-of-the-art performance on multiple cross-lingual benchmarks, demonstrating that large-scale multilingual architectures can effectively support code-switched and hybrid languages such as Hinglish.

Another approach explored in the study is **mBART** [17], which employs a denoising autoencoder model trained on monolingual data from 25 languages. Unlike models that focus solely on either the encoder or the decoder, mBART jointly trains both components, enabling it to learn more effective translation mappings. During training, noise is intentionally introduced through techniques such as token masking and sentence reordering, encouraging the model to reconstruct the original text. This training strategy enables mBART to perform effectively in both supervised and unsupervised settings, particularly when resources are limited. It also demonstrates strong performance in zero-shot translation

***Corresponding Author:** Prof. Samatha R. Swamy

tasks, highlighting its robustness in multilingual and code-switched applications.

The research further extends to low-resource languages, focusing on **Tatar** as a case study. One contribution presents a transliteration model that combines subword-level tokenization with language detection, specifically designed to convert Cyrillic Tatar mixed with Russian into a normalized form. Another contribution introduces a **Tatar Universal Dependencies treebank** that annotates code-switching at the morpheme level [13], capturing nuanced language shifts within individual words [18]. These efforts emphasize the importance of developing specialized linguistic resources and datasets to support underrepresented languages, particularly in multilingual and code-mixed environments.

The final section of the work evaluates **ChatGPT's** translation capabilities from English to Turkish. Using feedback from professional academic translators, the study assesses the fluency, semantic accuracy, and overall naturalness of the generated translations. The findings indicate that ChatGPT often produces translations comparable to those of human translators, particularly in terms of fluency. However, it occasionally struggles with idiomatic expressions and culturally specific meanings. These observations initiate broader discussions regarding the expanding role of AI in translation and its implications for the future of professional human translators in an increasingly technology-driven environment.

The increasing use of Hinglish in digital communication has created the need for sophisticated systems capable of handling its unique linguistic characteristics. This includes critical tasks such as hate speech detection and machine translation (MT). One notable study, *Fighting Hate Speech from a Bilingual Hinglish Speaker's Perspective: A Model and Translation-Based Approach*, proposes a novel framework that combines model-based techniques with translation methods to identify hate speech in Hinglish text. The study emphasizes the importance of understanding bilingual context and demonstrates how multilingual models such as **mBART** can improve both language identification and hate speech classification. The proposed framework also promotes the use of cross-lingual embeddings and highlights the importance of contextual information when interpreting code-mixed content.

Another significant contribution, presented in *CNLP-NITS-PP at MixMT 2022: Hinglish-English Code-Mixed Machine Translation*, introduces a system specifically designed for Hinglish-to-English translation. This approach leverages deep learning architectures, particularly sequence-to-sequence models, to address the linguistic challenges associated with code-mixed inputs, especially those originating from informal social media platforms. The system incorporates pre-trained models that are fine-tuned for Hinglish translation, resulting in notable improvements in BLEU scores and overall translation accuracy.

Expanding this research direction, the study *Hate Speech Recognition in Multilingual Hinglish Texts* [19], [23] focuses on identifying hate speech in Hinglish content, a challenging task due to the language's inherently mixed nature. Conventional NLP techniques often perform poorly in this domain; however, by employing both monolingual and cross-lingual embeddings, the researchers developed a model capable of capturing more nuanced multilingual representations of hate speech. This approach significantly outperformed earlier methods, demonstrating the effectiveness of embedding-based representations.

Building upon similar concepts, *MUCS@MixMT: IndicTrans-Based Machine Translation* [24] explores the application of **IndicTrans**, a translation model specifically developed for Indic languages, to Hinglish-to-English translation. The study demonstrates IndicTrans's ability to effectively process both formal and informal Hinglish text. By fine-tuning the model using Hinglish-specific datasets, the researchers achieved substantial improvements in translation quality, reinforcing the importance of domain-specific training.

Finally, the work *GUI [25] at MixMT 2022: English-Hinglish—A Machine Translation Approach for Code-Mixed Data* introduces a bidirectional translation framework [25] capable of performing both Hinglish-to-English and English-to-Hinglish translation. The framework incorporates preprocessing techniques such as vocabulary normalization and transliteration to improve the processing of code-mixed inputs. Evaluated using metrics including **BLEU** [20] and **Word Error Rate (WER)**, the system achieved significant improvements, although challenges remain because of the informal and continuously evolving nature of Hinglish, particularly on social media platforms. The authors emphasize the need to develop more diverse and representative datasets to further improve system performance.

Collectively, these studies make significant contributions to the advancement of Hinglish language processing, particularly in machine translation and hate speech detection. They highlight the importance of specialized datasets, multilingual transformer models, and cross-lingual embeddings [23] in addressing the linguistic and societal challenges associated with code-mixed languages. Despite these advances, continued research involving richer datasets and more robust translation models remains essential for overcoming the existing limitations in this rapidly evolving field.

MATERIALS AND METHODS

The proposed Hinglish-to-English translation system is designed using a combination of advanced Natural Language Processing (NLP) techniques, machine

***Corresponding Author:** Prof. Samatha R. Swamy

learning models, and structured processing pipelines to address the unique challenges associated with code-mixed language translation. Given the complex linguistic characteristics of Hinglish—including transliterations, phonetic inconsistencies, and mixed grammatical structures—a multi-pronged methodology consisting of four primary approaches was adopted.

Unified Multistage Pipeline

The proposed methodology decomposes the translation process into a sequence of manageable subtasks, enabling more accurate processing of Hinglish text.

- **Language Identification:** Each token in a sentence is classified as Hindi, English, or ambiguous using multilingual models such as **XLM-R** and **mBERT**.
- **Phonetic Transliteration:** Romanized Hindi words are converted into Devanagari script using an **LSTM-based Grapheme-to-Phoneme (G2P)** model.
- **Syntax Reordering:** Dependency parsing using BERT-based models reorganizes sentence structures into English-compliant syntax.
- **Neural Machine Translation (NMT):** The transformed sentence is translated using transformer-based models such as **mBART** or **IndicTrans**.
- **Post-Processing:** Rule-based grammatical correction improves the fluency, readability, and coherence of the translated output.

This pipeline is particularly suitable for formal applications such as chatbots, virtual assistants, and customer support systems.

Low-Latency Hybrid Model

Designed for real-time performance in resource-constrained environments, this approach combines rule-based methods, statistical machine translation, and deep learning techniques.

- **Lexicon-Based Word Substitution:** Frequently used Hinglish words are translated using predefined bilingual dictionaries.
- **Statistical Machine Translation (SMT):** Phrase-based statistical models handle short, informal expressions that are not covered by rule-based techniques.
- **Neural Machine Translation Integration:** Transformer models such as **IndicTrans** are employed to translate more complex sentence structures.
- **Scoring and Re-Ranking:** Candidate translations are evaluated using **BLEU** and **METEOR** scores, and the highest-scoring output is selected.

This hybrid framework is well suited for messaging applications and social media platforms because of its relatively low computational requirements while maintaining satisfactory translation accuracy.

Reinforcement Learning-Based Self-Improving Translator

To enable continuous learning and adaptability, a feedback-driven reinforcement learning approach was implemented.

- **Base Model:** A transformer-based Neural Machine Translation model (e.g., **mBART**) serves as the foundation.
- **Reinforcement Learning Fine-Tuning:** The translation model is continuously refined using user feedback, where correct translations receive positive reward signals.
- **Evaluation Metrics:** Translation quality improvements are measured using **BLEU** and **Translation Edit Rate (TER)** scores.
- **Self-Correction Mechanism:** User-provided corrections are incorporated into the training loop, allowing the model to learn personalized translation preferences over time.

This approach is particularly beneficial for enterprise applications, intelligent voice assistants, and translation systems requiring continuous adaptation in informal communication environments.

Hinglish-English Neural Machine Translation Model

The proposed system employs a transformer-based multilingual **mBART** model integrated with **IndicTrans** preprocessing modules to achieve accurate Hinglish-to-English translation.

- **Base Model:** A multilingual **mBART** model fine-tuned using the **MixMT** dataset together with a custom Hinglish-English parallel corpus.
- **IndicTrans Preprocessing:** IndicTrans normalization and tokenization modules are incorporated to improve the processing of code-mixed Hinglish text and named entities.
- **Fine-Tuning:** Model training is performed using the **Hugging Face Seq2Seq** framework, with hyperparameters optimized for multilingual translation tasks.
- **Evaluation Metrics:** Translation performance is comprehensively evaluated using **BLEU**, **Translation Edit Rate (TER)**, **Word Error Rate (WER)**, and overall translation accuracy.

This approach provides robust, adaptive, and context-aware translation for informal Hinglish text, making it suitable for conversational AI systems, social media analytics, educational platforms, and enterprise translation applications.

IMPLEMENTATION

The proposed Hinglish-to-English translation system directly processes Roman-script Hinglish without converting it into Devanagari, thereby aligning with real-world communication patterns commonly observed on social media platforms, chat applications, and conversational AI systems.

***Corresponding Author:** Prof. Samatha R. Swamy

The model was trained and evaluated using a combination of publicly available and custom datasets:

- **PHINC:** A parallel Hinglish-English corpus containing **13,736** sentence pairs.
- **CMU Hinglish Dog:** A dataset containing **9,962** Hinglish-English conversational sentence pairs.
- **MixMT 2022 Corpus:** A curated Hinglish-English dataset developed for machine translation.
- **Custom Dataset:** Hinglish-English sentence pairs collected from social media platforms and other informal text sources.

The preprocessing stage consisted of the following steps:

- **Normalization:** Correcting phonetic variations and inconsistent spellings (e.g., *"kya kr rahe ho"* → *"kya kar rahe ho"*).
- **Tokenization:** Segmenting Hinglish text using IndicTrans-compatible tokenizers.
- **Language Identification:** Identifying Hindi and English tokens using multilingual encoders such as **XLM-R**.

The translation model was developed using the **mBART** architecture, enhanced with **IndicTrans** preprocessing modules. Fine-tuning was performed using the **Hugging Face Seq2Seq** framework.

- **Fine-Tuning:** Conducted for **10 epochs** using the combined dataset.
- **Train-Test Splits:** Experiments were carried out using **60:40, 70:30, 80:20, and 90:10** data splits.
- **Hyperparameters:** Learning rate = $\mathbf{3 \times 10^{-5}}$, batch size = **32**, optimizer = **AdamW**.

Translation quality was evaluated using **BLEU**, **Word Error Rate (WER)**, and **Translation Edit Rate (TER)** metrics. Human evaluation based on fluency and adequacy was also conducted. The proposed model consistently outperformed the baseline **mBART** model and **Google Translate**, demonstrating its effectiveness in handling informal, code-mixed Hinglish text.

RESULTS

The translation model, based on **mBART** and fine-tuned using the **IndicTrans** framework, demonstrated a steady improvement in translation accuracy across successive training epochs. Accuracy increased from **0.0100** in **Epoch 1** to **0.1540** by **Epoch 20**. The most substantial improvement occurred during the first eight epochs, after which the rate of improvement gradually decreased, indicating model convergence.

The consistent improvement in translation accuracy can be attributed to **mBART's** multilingual contextual embeddings, which effectively capture both the syntactic and semantic characteristics of code-switched Hinglish, together with **IndicTrans's** capability to process transliterated text and Indic grammatical structures. The rapid improvement observed during the early training stages suggests that the model quickly learns common translation patterns, while the later epochs focus on refining translations involving less frequent linguistic structures.

These findings demonstrate that the combination of **mBART** and **IndicTrans** is well suited for Hinglish-to-English machine translation. Although extending training beyond **8–10 epochs** results in only marginal improvements, early stopping could significantly reduce computational cost without substantially affecting translation performance. Future work may focus on expanding the training dataset to include a broader range of informal expressions, regional variations, and slang to further enhance translation accuracy.

CONCLUSION AND FUTURE WORK

This work presented a Hinglish-to-English translation system built upon the **mBART** architecture and fine-tuned using the **IndicTrans** framework. The proposed methodology employed a unified multistage pipeline incorporating transliteration, text normalization, and syntactic reordering, enabling the model to effectively process code-switched text.

The integration of **mBART's** multilingual contextual embeddings with **IndicTrans's** specialized Indic language processing capabilities enabled the system to outperform baseline translation models in terms of **BLEU** and **Word Error Rate (WER)** scores. The progressive improvement in translation accuracy across training epochs demonstrates the model's ability to generalize effectively to both formal and informal Hinglish language constructs. Furthermore, the modular design of the proposed pipeline allows straightforward extension to other Indic code-switched languages, highlighting its adaptability for broader multilingual translation applications.

To further enhance context-aware translation, future research will explore advanced multilingual architectures such as **IndicTrans2** and fine-tuned **Large Language Models (LLMs)** within the existing translation pipeline. The system will also incorporate adaptive learning modules capable of personalizing translations according to individual user writing styles and preferences.

In addition, deployment optimization efforts will focus on lightweight mobile implementations and API-based integration with chatbots, educational platforms, and enterprise applications. These enhancements provide a strong foundation for developing scalable, real-time Hinglish translation systems with broad applicability in conversational AI, education, healthcare, and digital content moderation.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

***Corresponding Author:** Prof. Samatha R. Swamy

ACKNOWLEDGEMENTS

The authors sincerely thank the **RV Institute of Technology and Management (RVITM)** for their valuable support and encouragement throughout this research.

REFERENCES

1. Kumbhar, M., Thakre, K. Language Identification and Transliteration Approaches for Code-Mixed Text. *Journal of Engineering Science and Technology Review*, **17**(1), 2024.
2. Schäfer, H., Idrissi-Yaghir, A., Horn, P., Friedrich, C. Cross-Language Transfer of High-Quality Annotations: Combining Neural Machine Translation with Cross-Linguistic Span Alignment to Apply NER to Clinical Texts in a Low-Resource Language. *Proceedings of the 4th Clinical NLP Workshop*, pp. 53–62, 2022.
3. Kulkarni, M., Garera, N. Vernacular Search Query Translation with Unsupervised Domain Adaptation. *arXiv preprint arXiv:2208.03711*, 2022.
4. Litschko, R., Vulić, I., Ponzetto, S. P., Glavaš, G. On Cross-Lingual Retrieval with Multilingual Text Encoders. *Information Retrieval Journal*, **25**(2), 149–183, 2022.
5. Liu, Y., Ma, C., Ye, H., Schütze, H. TransliCo: A Contrastive Learning Framework to Address the Script Barrier in Multilingual Pretrained Language Models. *arXiv preprint arXiv:2401.06620*, 2024.
6. Jadhav, I., Kanade, A., Waghmare, V., Chandok, S. S., Jarali, A. Code-Mixed Hinglish to English Language Translation Framework. *Proceedings of the International Conference on Sustainable Computing, Data Communication Systems (ICSCDS)*, pp. 684–688, 2022.
7. Wu, Y., Qin, Y. Machine Translation of English Speech: Comparison of Multiple Algorithms. *Journal of Intelligent Systems*, **31**(1), 159–167, 2022.
8. Yehudai, A., Choshen, L., Fox, L., Abend, O. Reinforcement Learning with Large Action Spaces for Neural Machine Translation. *arXiv preprint arXiv:2210.03053*, 2022.
9. Bhuvaneswari, K., Varalakshmi, M. Efficient Incremental Training Using a Novel NMT-SMT Hybrid Framework for Translation of Low-Resource Languages. *Frontiers in Artificial Intelligence*, **7**, 1381290, 2024.
10. Kedar, S. V., Bhangale, S., Deokar, K., Deshmukh, S., Biradar, P. Translation: Code-Mixed Language (Hinglish) to English. *Mathematics, Statistics and Engineering Applications*, **71**(4), 159–167, 2022.
11. Raj, Y., Laddagiri, B. MATra: A Multilingual Attentive Transliteration System for Indian Scripts. *arXiv preprint arXiv:2208.10801*, 2022.
12. Yang, G., Chen, J., Lin, W., Byrne, B. Direct Preference Optimization for Neural Machine Translation with Minimum Bayes Risk Decoding. *arXiv preprint arXiv:2311.08380*, 2023.
13. Taguchi, C. Natural Language Processing for Lower-Resource Code-Switching Languages: Case Studies on Transliteration and Resource Building for Tatar, 2022.
14. Rao, A., D'Souza, N., Patel, D., Saravta, J. Development and Study of Hinglish-to-English Translation and Classification Techniques. *SSRN Preprint*, 4510958, 2023.
15. Kumar, A., Pratap, A., Singh, A. K. Generative Adversarial Neural Machine Translation for Phonetic Languages via Reinforcement Learning. *IEEE Transactions on Emerging Topics in Computational Intelligence*, **7**(1), 190–199, 2022.
16. Xue, L., *et al.* mT5: A Massively Multilingual Pre-Trained Text-to-Text Transformer. *arXiv preprint arXiv:2010.11934*, 2020.
17. Liu, Y., *et al.* Multilingual Denoising Pre-Training for Neural Machine Translation. *Transactions of the Association for Computational Linguistics*, **8**, 726–742, 2020.
18. Dalayli, F. Use of NLP Techniques in Translation by ChatGPT: A Case Study. *Proceedings of the ConTeNTS Workshop*, pp. 19–25, 2023.
19. Biradar, S., Saumya, S., Chauhan, A. Fighting Hate Speech from a Bilingual Hinglish Speaker's Perspective: A Transformer- and Translation-Based Approach. *Social Network Analysis and Mining*, **12**(1), 87, 2022.
20. Laskar, S. R., *et al.* CNLP-NITS-PP at MixMT 2022: Hinglish-English Code-Mixed Machine Translation. *Proceedings of the 7th Conference on Machine Translation (WMT)*, pp. 1158–1161, 2022.
21. Arora, G., Merugu, S., Sembium, V. CoMix: Guide Transformers to Code-Mix Using POS Structure and Phonetics. *Findings of the Association for Computational Linguistics: ACL*, pp. 7985–8002, 2023.
22. Hu, J., Hayashi, H., Cho, K., Neubig, G. DEEP: Denoising Entity Pre-Training for Neural Machine Translation. *arXiv preprint arXiv:2111.07393*, 2021.
23. Yadav, A. K., *et al.* Hate Speech Recognition in Multilingual Text: Hinglish Documents. *International Journal of Information Technology*, **15**(3), 1319–1331, 2023.
24. Hegde, A., Lakshmaiah, S. MUCS@MixMT: IndicTrans-Based Machine Translation for Hinglish Text. *Proceedings of the 7th Conference on Machine Translation (WMT)*, pp. 1131–1135, 2022.
25. Gahoi, A., *et al.* GUI at MixMT 2022: English-Hinglish—An MT Approach for Translation of Code-Mixed Data. *arXiv preprint arXiv:2210.12215*, 2022.

***Corresponding Author:** Prof. Samatha R. Swamy