

Brief Report

Not peer-reviewed version

---

# Assessing Transformers and Traditional Models for Spanish-English Code-Switched Hate Detection

---

[Steve Nwaiwu](#) \* and [Nipat Jongsawat](#)

Posted Date: 1 April 2025

doi: 10.20944/preprints202504.0052.v1

Keywords: Hate Speech Detection; Code-switching; Transformer Models; Multilingual NLP; Social Media; Spanglish



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

*Brief Report*

# Assessing Transformers and Traditional Models for Spanish-English Code-Switched Hate Detection

Steve Nwaiwu \* and Nipat Jongsawat

Data and Information Scienc, Faculty of Science and Technology, RMUTT, Pathum Thani, Thailand

\* Correspondence: chinonso\_s@mail.rmUTT.ac.th

**Abstract:** Hate speech detection research has mainly focused on monolingual contexts, with limited exploration of multilingual and code-switched languages that introduce distinct linguistic complexities. This study examines hate speech detection in code-switched Spanglish content from social media, comparing transformer-based models—XLM-RoBERTa, DistilBERT, Multilingual BERT, and mT5—with traditional machine learning approaches, including support vector machines, logistic regression, and multinomial naïve Bayes using TF-IDF features. The results indicate that XLM-RoBERTa achieves the highest performance, with 96.14 percent accuracy, 96.16 percent precision, 96.14 percent recall, and a 96.12 percent F1-score, demonstrating its superiority in detecting code-switched hate speech. Although traditional models, particularly SVM (94.03 percent accuracy), perform well, transformer-based approaches offer clear advantages in multilingual contexts.

**Keywords:** Hate Speech Detection; Code-switching; Transformer Models; Multilingual NLP; Social Media; Spanglish

## 1. Introduction

### 1.1. Background

Code-switching, the alternation between two or more languages within a single discourse, is prevalent in multilingual communities, particularly in online and social media interactions. This linguistic phenomenon introduces structural, semantic, and contextual complexities that pose significant challenges to natural language processing (NLP) models [1]. Hate speech detection, a critical NLP task, has been extensively studied in monolingual settings, achieving high classification accuracy through traditional machine learning methods and advanced deep learning models [2,3]. However, these models exhibit significant performance degradation when applied to code-switched text, which is characterized by abrupt linguistic shifts, inconsistent grammar, and culturally embedded expressions [4].

Transformer-based models, such as BERT and its multilingual variants (e.g., mBERT and XLM-RoBERTa), have demonstrated considerable effectiveness in multilingual NLP tasks. However, their potential in accurately detecting hate speech in code-switched contexts remains underexplored. Code-switching introduces unique linguistic dependencies and contextual nuances that conventional NLP pipelines struggle to capture, leading to reduced accuracy and ineffective content moderation. To address this gap, the present study investigates hate speech detection in Spanglish, a fluid blend of Spanish and English widely used across digital platforms, with the aim of evaluating and enhancing the performance of transformer-based models in these complex linguistic settings.

### 1.2. Research Problem

Most existing hate speech detection frameworks are designed for monolingual datasets, limiting their applicability in code-switched environments that involve lexical borrowing, grammatical complexity, and nuanced semantic shifts. A major challenge in advancing code-switched hate speech detection is the scarcity of specialized datasets and the lack of fine-tuning methodologies optimized

for multilingual data. Traditional approaches, including Support Vector Machines (SVM), Logistic Regression, and Naïve Bayes classifiers with TF-IDF features, have shown success in monolingual scenarios but struggle to generalize in multilingual and code-switched settings, where lexical boundaries and semantic polarity shift rapidly [4].

Although multilingual transformer models, such as XLM-RoBERTa, are trained on large-scale multilingual datasets, their efficiency and precision in distinguishing between hateful and nonhateful content in code-switched text require a thorough empirical assessment [14]. The key challenge lies in their ability to capture context-sensitive nuances that arise from simultaneous multilingual interactions within individual sentences. This study evaluates the effectiveness of fine-tuning transformer models and compares their performance with traditional machine learning models, analyzing their strengths and limitations across different data resource scenarios [15].

### 1.3. Research Objectives

To address the identified research challenges, this study pursues the following objectives:

- **Fine-tune transformer-based models** (XLM-RoBERTa, DistilBERT, mBERT, mT5) to improve accuracy in hate speech detection within Spanglish, leveraging multilingual contextual embeddings.
- **Benchmark the performance of transformer-based models** against traditional machine learning classifiers (SVM, Logistic Regression, and Multinomial Naïve Bayes) using TF-IDF features, providing a comprehensive analysis of their effectiveness in handling the linguistic complexities of code-switched text.
- **Explore low-resource fine-tuning strategies**, including weak supervision with multilingual lexicons.

Through these objectives, this study aims to enhance the capability and adaptability of NLP models for hate speech detection in multilingual and code-switched contexts, ultimately contributing to more robust automated moderation and improved regulation of online discourse across diverse digital platforms.

## 2. Literature Review

### 2.1. Hate Speech Detection in Monolingual Contexts

Hate speech detection has been a central topic in Natural Language Processing (NLP), focusing mainly on monolingual datasets, particularly in English, due to the availability of extensive annotated corpora [2,3]. Early approaches relied on statistical and machine learning classifiers, such as logistic regression, support vector machines (SVM), and multinomial naïve Bayes, often combined with feature engineering techniques like term frequency-inverse document frequency (TF-IDF) and n-grams. Davidson et al. [2] demonstrated that such feature-based models effectively distinguished hate speech from neutral or offensive content.

With the advancement of deep learning, neural network architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) significantly improved detection accuracy. A major breakthrough emerged with the introduction of the transformer-based BERT model [5], which enabled more effective contextual word embeddings. Subsequent improvements included models such as HateBERT, a fine-tuned version of BERT explicitly trained on hate speech datasets, achieving state-of-the-art performance [6]. Nwaiwu et al. [7] further demonstrated the application of fine-tuned transformer models in political discourse, significantly enhancing the accuracy of hate speech detection. However, despite these advances, most models remain confined to monolingual datasets, leaving their effectiveness in multilingual and code-switched contexts largely unexamined.

### 2.2. NLP Challenges in Code-Switched Contexts

Code switching, the alternation of languages within a single discourse, introduces significant challenges to NLP tasks due to its structural, semantic, and grammatical complexities [1,8]. Traditional NLP pipelines, predominantly designed for monolingual inputs, struggle to handle inconsistencies

in grammar, ambiguous semantics, and inter-sentential language shifts, leading to performance degradation when applied to multilingual and code-switched scenarios.

Various NLP tasks, including sentiment analysis, named entity recognition, and hate speech detection, exhibit unique difficulties in handling code-switched text [9]. For instance, switching languages mid-sentence can alter sentiment polarity or entity interpretation, necessitating models with enhanced multilingual contextual understanding.

Recent advances in multilingual NLP have introduced transformer-based cross-lingual models, such as XLM-RoBERTa, designed to take advantage of shared linguistic representations in multiple languages [10]. Although these models have shown promising results in cross-lingual benchmarks, their application to hate speech detection in code-switched contexts remains underexplored, particularly with respect to their robustness and effectiveness in resource-constrained, real-world environments.

### 2.3. Low-Resource NLP and Code-Switching

Low-resource NLP contexts, characterized by limited annotated data, have traditionally relied on machine learning models such as SVM, logistic regression, and multinomial naïve Bayes, which use simpler feature-based techniques such as TF-IDF and lexicon-based sentiment dictionaries [16]. Although these models often exhibit robustness in data-scarce environments, they fail to capture deep semantic relationships and complex context-dependent interactions, limitations that are particularly pronounced in code-switched text [17].

Multilingual transformer models, trained on extensive multilingual corpora, have demonstrated potential in low-resource NLP tasks by transferring contextual knowledge learned between languages [11]. However, their efficacy in accurately capturing intra-sentential language shifts inherent to code-switching has not been extensively validated.

Hybrid approaches that combine deep learning architectures with traditional linguistic resources, such as multilingual hate speech lexicons, have emerged as promising alternatives. Techniques such as weak supervision using lexicons, transfer learning, domain adaptation, and pseudolabeling offer potential solutions to the limitations of traditional or purely data-driven linguistic methodologies [9,11]. Therefore, comparative evaluations of traditional models and transformer-based architectures are essential to identify optimal strategies for handling code-switched and low-resource NLP environments.

### 2.4. Summary of Literature and Research Gap

The existing literature highlights substantial progress in the detection of hate speech within monolingual contexts, particularly with the adoption of transformer-based models such as HateBERT and XLM-RoBERTa. However, code-switching introduces unique linguistic challenges that remain inadequately addressed by current models and methodologies [18]. Although multilingual transformers have shown promising results in cross-lingual NLP tasks, their performance in detecting hate speech within code-switched languages, specifically Spanglish, remains underexplored.

To address this gap, the present study systematically evaluates the effectiveness of transformer-based models against traditional machine learning benchmarks in detecting hate speech within code-switched Spanglish social media content. Through comparative analysis, this research provides critical insights into the applicability and performance of multilingual transformers in real-world multilingual settings, laying the foundation for future advancements in NLP models tailored specifically for multilingual and code-switched environments.

## 3. Methodology

### 3.1. Dataset Description

This study used a publicly available dataset from the SemEval2020 challenge [12], specifically designed to detect hate speech in Spanglish—a code-switched mixture of Spanish and English commonly found on social media platforms. The dataset is structured as shown in Table 1.

Table 1. Dataset Composition

Dataset Split	Number of Samples
Training Set	11,999
Validation Set	3,000
Test Set	6,500

To enhance labeling accuracy through weak supervision, two comprehensive hate lexicons were incorporated, sourced from HateBase [13]. The details of the lexicons are presented in Table 2.

Table 2. Hate Lexicon Details

Lexicon	Number of Terms
English Hate Lexicon	5,963
Spanish Hate Lexicon	3,354

3.2. Data Preprocessing

The preprocessing pipeline consisted of multiple steps to ensure high-quality input data for model training. The text cleaning process included:

- **Lowercasing:** Converting all text to lowercase for consistency.
- **Noise Removal:** Eliminating punctuation marks, URLs, special characters, numbers, and user mentions (@user).
- **Whitespace Normalization:** Standardization of spacing within and between sentences.

3.2.1. Tokenization

The cleaned text was tokenized using transformer-specific tokenizers tailored for each model, including XLM-RoBERTa, DistilBERT, Multilingual BERT (mBERT), and multilingual T5 (mT5). Tokenization was essential in breaking down text into sub-word units, allowing transformer models to better capture linguistic nuances.

For traditional machine learning models, the cleaned textual data was converted into numerical representations using **Term Frequency-Inverse Document Frequency (TF-IDF)**. The TF-IDF score for a term  $t$  in document  $d$  is mathematically defined as:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t) \tag{1}$$

where:

$$\text{TF}(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \tag{2}$$

$$\text{IDF}(t) = \log\left(\frac{N}{n_t + 1}\right) \tag{3}$$

where:

- $f_{t,d}$  represents the frequency of term  $t$  in document  $d$ .
- $N$  is the total number of documents in the corpus.
- $n_t$  denotes the number of documents that contain the term  $t$ .

3.3. Models and Experiments

3.3.1. Transformer-Based Deep Learning Models

Four multilingual transformer-based models were employed:

- **XLM-RoBERTa:** A cross-lingual transformer pre-trained on multilingual data.
- **DistilBERT:** A lightweight and distilled version of BERT optimized for efficiency.
- **Multilingual BERT (mBERT):** A BERT model pre-trained on multiple languages.

- **Multilingual T5 (mT5):** A transformer encoder-decoder model designed for text-to-text tasks.

The training process used the following hyperparameters:

- **Learning Rate:**  $2 \times 10^{-5}$
- **Batch Size:** 16 samples for training and validation.
- **Epochs:** 5
- **Weight Decay:** 0.01 (to reduce overfitting).
- **Optimizer:** AdamW, known for stabilizing transformer training.

The validation loss was monitored after each epoch, and the best performing model was selected based on the lowest validation loss. Checkpointing was used to save the two best performing models during training, with the top performing model automatically loaded after training.

### 3.3.2. Traditional Machine Learning Models

To establish a baseline, three traditional classifiers were evaluated:

- **Logistic Regression:** Efficient and interpretable model.
- **Support Vector Machines (SVM):** Effective in high-dimensional text data.
- **Multinomial Naïve Bayes:** Probabilistic classifier commonly used for text classification.

All classifiers used TF-IDF vectors as input, providing a comparative benchmark against transformer-based models.

### 3.4. Evaluation Metrics

To assess model performance, multiple evaluation metrics were used:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

where:

- **TP (True Positive):** Correctly identified hate speech instances.
- **TN (True Negative):** Correctly identified non-hate speech instances.
- **FP (False Positive):** Incorrectly classified non-hate speech as hate speech.
- **FN (False Negative):** Hate speech instances misclassified as non-hate.

Additionally, classification reports detailing the performance of individual classes (hate and non-hate) were generated to provide deeper insights into model behavior.

### 3.5. Computational Resources

All experiments were carried out on GPU-accelerated infrastructure to optimize computational efficiency and reduce training times. Runtime metrics such as:

- **Training Duration**
- **Speed (Samples per Second)**
- **Total Floating-Point Operations (FLOPs)**

were systematically recorded to evaluate computational demands and model efficiency.

3.6. Ethical Considerations

Due to the sensitive nature of hate speech detection, ethical considerations were rigorously observed. The dataset’s anonymity and user privacy were strictly maintained. Data handling procedures complied with ethical research guidelines, ensuring responsible model development and deployment.

4. Results and Discussion

This section presents a comparative analysis of transformer-based deep learning models against traditional machine learning (ML) benchmarks. The evaluation focuses on four key metrics: accuracy, precision, recall, and F1-score. Traditional ML models serve as foundational benchmarks for assessing the improvements achieved by advanced transformer-based approaches.

4.1. Comprehensive Performance Comparison

The results of both transformer-based and traditional ML models are systematically summarized in Table 3. The highest scores in each metric are highlighted in **bold** for clarity.

Table 3. Comparative Performance of Transformer-Based and Traditional ML Models

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression (Benchmark)	0.9065	<b>0.9920</b>	0.7698	0.8669
SVM (Benchmark)	0.9403	0.9995	0.8495	0.9185
Multinomial Naïve Bayes (Benchmark)	0.9074	0.9616	0.7978	0.8721
<b>XLM-RoBERTa</b>	<b>0.9614</b>	0.9616	<b>0.9614</b>	<b>0.9612</b>
DistilBERT	0.9423	0.9432	0.9423	0.9419
mBERT	0.9594	0.9599	0.9594	0.9592
mT5	0.6305	0.6146	0.6305	0.5942

As shown in Table 3, the transformer-based **XLM-RoBERTa** model emerged as the top-performing approach, achieving the highest accuracy (96.14%), recall (96.14%), and F1-score (96.12%). This demonstrates its superior ability to effectively classify code-switched hate speech in Spanglish.

The **mBERT** model exhibited strong competitive performance, following closely behind XLM-RoBERTa, indicating its suitability as an alternative. Meanwhile, **DistilBERT** provided a balance between computational efficiency and accuracy, making it a viable choice for scenarios with limited computational resources.

Among traditional ML models, the **Support Vector Machine (SVM)** model showed robust performance with notably high precision (99.95%) and substantial accuracy (94.03%), underscoring its utility in resource-limited scenarios, albeit with slightly lower performance compared to transformer models.

Interestingly, the **mT5** model significantly underperformed, recording the lowest accuracy (63.05%), precision (61.46%), recall (63.05%), and F1-score (59.42%). This suggests that generative transformer models, such as mT5, may be less suitable for hate speech detection tasks involving complex code-switching phenomena.

4.2. Discussion of Results

The superior results achieved by **XLM-RoBERTa** can be attributed to its robust pretraining strategy on large multilingual datasets, enabling it to capture nuanced linguistic patterns characteristic of code-switched language environments. Its enhanced contextual awareness and ability to distinguish subtle semantic differences significantly contribute to its performance in hate speech classification across multilingual content.

Traditional ML methods, while effective, demonstrated limitations due to their reliance on hand-crafted feature representations (e.g., TF-IDF), which lack the contextual depth captured by transformer-based embeddings. However, their computational efficiency and speed make them advantageous in real-time or resource-constrained applications.

4.3. Computational Efficiency

In terms of computational efficiency, **DistilBERT** provided an optimal trade-off, offering strong performance while consuming fewer computational resources compared to larger models such as XLM-RoBERTa and mBERT. Traditional models, such as logistic regression and multinomial naïve Bayes, proved exceptionally efficient, but exhibited constraints in accuracy and general linguistic understanding.

Table 4 highlights the efficiency of each model. **Logistic Regression** and **Multinomial Naïve Bayes** were the fastest models in terms of training speed and inference, making them ideal for low-latency applications. In contrast, **XLM-RoBERTa** and **mBERT**, despite their superior accuracy, required significantly longer training times and lower inference speeds.

Table 4. Computational Efficiency of Models

Model	Training Time (min)	Inference Speed (sec)
Logistic Regression	1.2	3100
SVM	3.5	1200
Multinomial Naïve Bayes	1.4	2800
XLM-RoBERTa	52.7	120
DistilBERT	27.8	180
mBERT	49.3	130
mT5	65.2	95

In general, the findings indicate that transformer-based models, particularly **XLM-RoBERTa**, offer superior capabilities to detect hate speech within code-switched Spanglish text. These results emphasize the importance of exploiting context-rich transformer architectures for multilingual NLP applications. The insights gained from this study lay a solid foundation for future research and practical implementations in multilingual content moderation.

5. Conclusion and Future Work

5.1. Summary of Findings

This study comprehensively investigated the task of hate speech detection within Spanglish—a dynamic blend of Spanish and English frequently used in social media—by conducting a comparative analysis between transformer-based deep learning models (XLM-RoBERTa, DistilBERT, mBERT, mT5) and traditional machine learning classifiers (SVM, Logistic Regression, and Multinomial Naïve Bayes) utilizing TF-IDF features. The key findings of this research are as follows:

- Transformer-based models significantly outperformed traditional classifiers, with **XLM-RoBERTa** emerging as the best-performing model, achieving the highest accuracy (96.14%), precision (96.16%), recall (96.14%), and F1-score (96.12%). This underscores the superior capability of transformer architectures in capturing the linguistic complexities inherent in code-switched text.
- Among traditional classifiers, the **Support Vector Machine (SVM)** model demonstrated strong performance, with competitive accuracy (94.03%) and precision (99.95%), proving its effectiveness in scenarios with limited computational resources.
- The **mT5** model exhibited significantly lower performance (accuracy: 63.05%), indicating the limitations of certain transformer architectures, particularly generative models, in handling code-switched language tasks.
- The error analysis revealed persistent linguistic challenges, including slang usage, idiomatic expressions, negations, and semantic ambiguities in code-switched text. These complexities pose

significant challenges for existing NLP models, highlighting areas requiring further refinement to enhance model robustness in multilingual environments.

### 5.2. Contributions and Implications

This study makes several key contributions to the field of multilingual hate speech detection:

- Provides in-depth evaluation of transformer-based and traditional ML methods, offering empirical evidence of the superiority of transformer models in handling the intricacies of code-switched text.
- The research introduces an enriched dataset enhanced through weak supervision by integrating extensive multilingual lexicons, thereby facilitating further research in multilingual and code-switched NLP applications.
- The study identifies critical linguistic challenges that affect model accuracy, particularly the detection of slang and implicit hate speech expressions, emphasizing the need for targeted enhancements in the design of the NLP model.
- The findings have practical implications for **automated content moderation, online safety, and inclusive NLP applications**, particularly in diverse and multilingual digital communities. Improved hate speech detection models can contribute to safer and more responsible online discourse by mitigating the spread of harmful content.

### 5.3. Future Work

Although this study provides valuable information, several avenues for future research can further enhance hate speech detection in multilingual and code-switched environments.

- **Data Augmentation Techniques:** Expanding existing datasets through synthetic data generation, paraphrasing, and translation-based augmentation can help address data scarcity issues and improve model generalization.
- **Few-Shot and Zero-Shot Learning Approaches:** Developing models that can achieve high classification accuracy with minimal labeled data is crucial in resource-constrained multilingual settings. Exploring meta-learning and contrastive learning strategies can facilitate effective learning with limited training examples.
- **Hybrid Models and Fusion Strategies:** Integrating linguistic heuristics, lexicon-based methods, and transformer embeddings can improve the detection of implicit and nuanced hate speech, particularly in informal and context-dependent conversations.
- **Domain Adaptation and Transfer Learning:** Investigating the adaptability of transformer-based models across various digital platforms, such as social media, online forums, and conversational agents, will ensure consistent and reliable hate speech detection in real-world applications.
- **Ethical AI and Bias Mitigation:** Future studies should explore techniques to reduce biases in hate speech detection models to ensure fairness between different demographic and linguistic groups, minimizing the risk of unintended model biases and false positives.

Addressing these research directions will contribute significantly to the development of more robust, accurate, and context-aware NLP solutions. Enhancing hate speech detection in code-switched multilingual contexts is crucial for fostering safer, more inclusive digital communication environments, ultimately promoting ethical and responsible AI applications in online discourse.

**Acknowledgments:** The authors would like to thank the Dean of the Faculty of Science and Technology, RMUTT, for their support and contributions to this research.

## References

1. T. Solorio, Y. Liu, A. Rios, and F. L. Cruz, "Part-of-speech tagging for English-Spanish code-switched text," *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1051–1060, 2014. Available: <https://aclanthology.org/D14-1112>

2. T. Davidson, D. Warmesley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language," *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*, pp. 512–515, 2017. Available: <https://ojs.aaai.org/index.php/ICWSM/article/view/14955>
3. Z. Waseem and D. Hovy, "Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter," *Proceedings of the NAACL Student Research Workshop*, pp. 88–93, 2016. Available: <https://aclanthology.org/N16-2013>
4. G. Molina, A. Cabrera, and T. Solorio, "Overview for the second shared task on language identification in code-switched data," *Proceedings of the Second Workshop on Computational Approaches to Code Switching*, pp. 40–49, 2016. Available: <https://aclanthology.org/W16-5806>
5. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pp. 4171–4186, 2019. Available: <https://aclanthology.org/N19-1423>
6. T. Caselli, V. Basile, J. Mitrović, I. Kartoziya, and M. Granitzer, "HateBERT: Retraining BERT for abusive language detection in English," *arXiv preprint arXiv:2010.12472*, 2020. Available: <https://arxiv.org/abs/2010.12472>
7. S. Nwaiwu, N. Jongsawat, A. Tungasthan, and J. Thaloey, "Fine-tuned BERT model for hate speech detection in political discourse," *2024 22nd International Conference on ICT and Knowledge Engineering (ICT&KE)*, pp. 1–8, 2024. <https://doi.org/10.1109/ICTKE62841.2024.10787162>
8. A. Pratapa, S. Bhat, K. Bali, and M. Choudhury, "Language modeling for code-mixing: The role of linguistic theory-based synthetic data," *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1543–1553, 2018. Available: <https://aclanthology.org/P18-1143>
9. G. I. Winata, A. Madotto, and P. Fung, "Code-switching language modeling using syntax-aware multi-task learning," *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 3786–3798, 2021. Available: <https://aclanthology.org/2021.acl-long.293>
10. A. Conneau *et al.*, "Unsupervised cross-lingual representation learning at scale," *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 8440–8451, 2020. Available: <https://aclanthology.org/2020.acl-main.747>
11. T. Pires, E. Schlinger, and D. Garrette, "How multilingual is multilingual BERT?" *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 4996–5001, 2019. Available: <https://aclanthology.org/P19-1493>
12. G. Aguilar, S. Kar, T. Solorio, and G. Molina, "LinCE: A centralized benchmark for linguistic code-switching evaluation," *Proceedings of the 12th Language Resources and Evaluation Conference (LREC)*, pp. 1803–1813, 2020. Available: <https://aclanthology.org/2020.lrec-1.224>
13. Hatebase. (2019) "The world's largest structured repository of regionalized hate speech." [Online]. Available: <https://hatebase.org>. [Accessed: Mar. 16, 2025].
14. T. Tița and A. Zubiaga, "Cross-lingual Hate Speech Detection using Transformer Models," *arXiv preprint arXiv:2111.00981*, Nov. 2021. Available: <https://arxiv.org/abs/2111.00981>
15. Y. Zhang, P.-H.C. Chen, K. Gadepalli, *et al.*, "Transformer versus traditional natural language processing models for radiology report classification: A multi-institutional study," *Radiology: Artificial Intelligence*, vol. 5, no. 5, pp. e220207, 2023. DOI: [10.1148/ryai.220207](https://doi.org/10.1148/ryai.220207)
16. M. Das, K. Selvakumar, and P. J. A. Alphonse, "A Comparative Study on TF-IDF Feature Weighting Method and its Analysis using Unstructured Dataset," *arXiv preprint arXiv:2308.04037*, Aug. 2023. Available: <https://arxiv.org/abs/2308.04037>
17. M. H. Shakeel and A. Karim, "Adapting Deep Learning for Sentiment Classification of Code-Switched Informal Short Text," *arXiv preprint arXiv:2001.01047*, Jan. 2020. Available: <https://arxiv.org/abs/2001.01047>
18. A. S. Dogruöz and Ç. Çöltekin, "Representativeness as a Forgotten Lesson for Multilingual and Code-Switching NLP," *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 5190–5204, 2023. Available: <https://aclanthology.org/2023.findings-emnlp.382.pdf>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.