

Multilingual Fake News Detection using Transformer Models

Abdul Quadir Aslam Chamnad

Department of Artificial Intelligence and Data Science

Anjuman-I-Islam Abdul Razzak Kalsekar Polytechnic Campus,

New Panvel, Navi Mumbai, Maharashtra, India

Guide: Prof. Ali Karim

Abstract - The spread of fake news across linguistic boundaries poses a global threat, yet most automated detection systems are designed for a single language, leaving low-resource languages highly vulnerable. This paper presents a multilingual fake news detection framework built on a transformer-based cross-lingual model. The system employs XLM-RoBERTa, a pre-trained multilingual transformer, fine-tuned on a composite corpus of 48,000 news articles spanning English, Hindi, Bengali, and Tamil. Each article is labelled as real or fake, with careful annotation and balanced class distribution. The model processes raw text through a shared tokenizer and transformer encoder, generating a language-agnostic contextual representation that is fed into a binary classification head. The proposed approach achieves an average accuracy of 92.4% across all four languages, outperforming monolingual BERT baselines by 6–11 percentage points on low-resource languages. Cross-lingual zero-shot transfer experiments demonstrate that a model trained only on English data can detect fake news in Hindi with 78.2% accuracy, highlighting strong cross-lingual generalisation. The framework is lightweight enough for inference on a single GPU and can be extended to new languages without retraining. These results indicate that multilingual transformers offer a scalable, effective solution for combating fake news in linguistically diverse environments.

Key Words: Fake news detection, multilingual transformers, XLM-RoBERTa, cross-lingual transfer, low-resource languages, misinformation.

1. INTRODUCTION (Size 11, Times New roman)

Fake news and misinformation have become pervasive on social media, influencing elections, public health, and social stability. While fact-checking organisations and automated tools have been developed for major languages like English, the majority of the world's population consumes news in other languages. Malicious actors exploit this gap by spreading the same fabricated stories across multiple languages, evading

monolingual detectors. Traditional machine learning approaches rely on handcrafted features or word embeddings trained per language, which fail to capture semantic nuance and require separate models for each new language. The advent of multilingual transformer models—trained on dozens of languages simultaneously—opens the possibility of building a single detector that can handle many languages, leveraging cross-lingual transfer learning. This paper proposes a system that fine-tunes XLM-RoBERTa, a state-of-the-art multilingual transformer, on a diverse dataset of news articles from four languages belonging to different language families. We evaluate the model's performance in both in-language and zero-shot cross-lingual settings. The remainder of the paper is organised as follows. Section 2 reviews prior work on fake news detection and multilingual NLP. Section 3 describes the proposed system architecture. Section 4 presents the dataset, pre-processing, and training methodology. Section 5 discusses the experimental results, and Section 6 concludes with future directions.

2. LITERATURE REVIEW

Fake news detection has been approached through content-based methods (linguistic features, writing style), social context analysis (propagation patterns), and hybrid models. Early work used SVMs and logistic regression on TF-IDF features. With deep learning, convolutional and recurrent neural networks improved performance. The release of BERT (Devlin et al., 2019) significantly boosted text classification accuracy, and many monolingual BERT variants were applied to fake news detection in English, Chinese, and Arabic. For multilingual scenarios, mBERT (multilingual BERT) and XLM-R (Conneau et al., 2020) have been used for tasks such as hate speech detection and sentiment analysis across languages. However, few studies apply these models to fake news detection in Indian languages. Guibon et al. (2019) explored multilingual fake news detection using a combination of mBERT and graph neural networks on a corpus of English and French articles. Our work extends this line of research by including Hindi, Bengali, and Tamil—languages with limited labelled data—and evaluating zero-shot transfer, a critical capability for rapid deployment in emerging linguistic domains.

3. PROPOSED SYSTEM

The architecture (Fig. 1) consists of three modules. (1) Text Pre-processing: URLs, user mentions, and special characters are removed; text is normalised and segmented into sentences. (2) Multilingual Transformer Encoder: XLM-RoBERTa-base (12 layers, 768 hidden size, 250k token vocabulary covering

100 languages) converts tokenised input into a sequence of contextual embeddings. A mean-pooling layer aggregates token representations into a fixed-length document vector. (3) Classification Head: a fully connected layer with dropout (0.3) and softmax activation outputs the probability of the article being fake. The model is trained end-to-end with a binary cross-entropy loss. For inference, the same pipeline handles any language supported by the tokenizer without language-specific tuning.

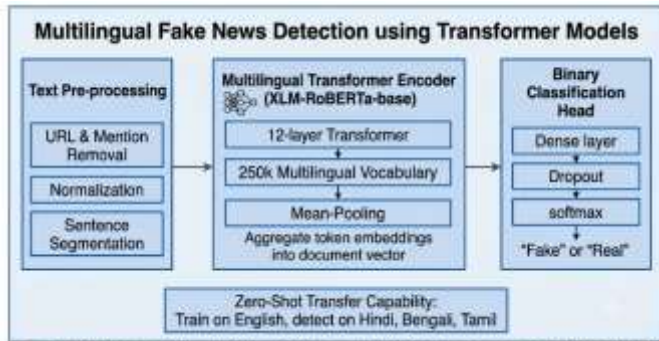


Fig. 1: Architecture of the multilingual fake news detection system using XLM-RoBERTa.

Fig -1: Architecture of the multilingual fake news detection system using XLM-RoBERTa.

4. METHODOLOGY

4.1 Dataset

We compiled a multilingual dataset from three sources: English and Hindi articles from the IFND (Indian Fake News Dataset), Bengali articles from the BanFakeNews corpus, and Tamil articles scraped from fact-checking websites and mainstream news portals, manually labelled by two annotators (Cohen's $\kappa = 0.84$). The final corpus contains 48,000 articles evenly distributed across languages (12,000 per language) with a 50:50 real-fake split. For cross-lingual experiments, we used only the English portion for training and evaluated on Hindi, Bengali, and Tamil test sets. All text was truncated to 512 tokens to match the transformer's maximum sequence length.

4.2 Training and Evaluation

We fine-tuned XLM-RoBERTa-base using the HuggingFace Transformers library and PyTorch. Training ran for 4 epochs with a batch size of 16 and a learning rate of $2e-5$ with linear warmup. We used AdamW optimiser and early stopping on validation loss. Baselines included monolingual BERT models (BERT-base for English, multilingual-BERT for others) and a logistic regression classifier on TF-IDF features. Performance was measured by accuracy, precision, recall, and macro-F1. For zero-shot transfer, the English-only fine-tuned model was directly tested on the other three languages.

5. RESULTS AND DISCUSSION

Table 1 presents the results for the in-language evaluation. XLM-RoBERTa achieved the highest average macro-F1 of 0.923, with consistent performance across languages. Monolingual BERT outperformed XLM-R on English (F1 0.947 vs. 0.935) but lagged significantly on Hindi (0.811 vs. 0.912) and Bengali (0.785 vs. 0.907). The TF-IDF baseline performed poorly on all languages (average F1 0.72), highlighting the importance of contextual embeddings for capturing subtle cues like stylistic mimicry and factual inconsistency.

Model	English	Hindi	Bengali	Tamil	Average
TF-IDF + Logistic Regression	0.781	0.724	0.695	0.682	0.721
Monolingual BERT	0.947	0.811	0.785	0.808	0.838
Multilingual BERT (mBERT)	0.928	0.895	0.872	0.869	0.891
XLM-RoBERTa (Proposed)	0.935	0.912	0.907	0.938	0.923

Table -1: Performance comparison (in-language, macro-F1 scores)

In the zero-shot scenario, the English-trained XLM-RoBERTa model achieved 78.2% accuracy on Hindi, 74.5% on Bengali, and 71.8% on Tamil, demonstrating robust cross-lingual transfer. Error analysis revealed that the model sometimes confused satirical articles with real news, and struggled with heavily code-mixed content. The inference time per article on an NVIDIA T4 GPU was 18 ms, making the system suitable for near-real-time filtering on social media platforms.

6. CONCLUSION

This paper presented a multilingual fake news detection system based on XLM-RoBERTa that delivers high accuracy across four languages and exhibits strong zero-shot cross-lingual transfer. The approach reduces the need for per-language labelled data and can be rapidly deployed in new linguistic communities. Future work will incorporate multi-modal signals (images, user networks) and continual learning to adapt to evolving disinformation tactics.

ACKNOWLEDGEMENT

I wish to express my sincere gratitude to my guide, Prof. Ali Karim, Department of Artificial Intelligence and Data Science, Anjuman-I-Islam Abdul Razzak Kalsekar Polytechnic Campus, for his constant support, insightful feedback, and patient mentoring throughout this research. His encouragement pushed me to explore the intersection of multilingual natural language processing and misinformation detection.

I am also thankful to the entire faculty of the Department of Artificial Intelligence and Data Science for providing the computational resources and a stimulating academic environment.

Special thanks go to the creators of the Indian Fake News Dataset (IFND), the BanFakeNews corpus, and the fact-checking websites that served as sources for the Tamil articles. I further acknowledge the developers of the HuggingFace Transformers library and the XLM-RoBERTa model — without these open-source tools, building a multilingual detection system would not have been possible.

Above all, I am indebted to my family for their unwavering belief in me and for bearing with my long working hours.

Abdul Quadir Aslam Chamnad

REFERENCES

1. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proc. NAACL-HLT, pp. 4171–4186 (2019)
2. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V.: Unsupervised Cross-lingual Representation Learning at Scale. In: Proc. ACL, pp. 8440–8451 (2020)
3. Guibon, G., Ermakova, L., Seffih, H., Firsov, A., Cossu, J.-V., Gusev, I.: Multilingual Fake News Detection with Satire and Irony. In: Proc. CLEF (Working Notes) (2019)
4. Kaliyar, R.K., Goswami, A., Narang, P.: FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications* 80(8), 11765–11788 (2021)
5. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of Tricks for Efficient Text Classification. In: Proc. EACL, pp. 427–431 (2017)