On Enhancing Code-Mixed Sentiment and Emotion Classification Using FNet and FastFormer

Anuj Kumar[1], Amit Pandey[1], Satyadev Ahlawat[2] and Yamuna Prasad[1]

[1] *Department of Computer Science & Engineering, Indian Institute of Technology Jammu, India*

[2] *Department of Electrical Engineering, Indian Institute of Technology Jammu, India*

{*anuj, 2022pis0027, satyadev.ahlawat, yamuna.prasad*}*@iitjammu.ac.in*

Keywords: Transformer, Cross-Lingual, Code-Mixed, Classification, FastFormer, FNet.

Abstract: Code-mixing, the blending of multiple languages within a communication, is becoming increasingly common on social media. If left unchecked for sentiment analysis, this trend can lead to hate speech or violence, emphasizing the need for advanced techniques to interpret emotions and sentiments in code-mixed languages accurately. Current research has mainly focused on code-mixed text involving a limited number of languages. However, these methods often yield suboptimal results due to inadequate feature extraction by existing learning models. Additionally, achieving high accuracy and extracting meaningful features from code-mixed text remains a significant challenge. To address this, we propose two transformer-based feature extraction methods for sentiment and emotion classification in code-mixed text. The first method integrates the Fourier transform into the transformer-based cross-lingual language model, XLM-Roberta, by incorporating the encoder layers of Fourier Net (FNet). This Fourier encoder layer applies a Fourier transform to the final output vector of hidden states, enabling the model to capture complex patterns more effectively. The second method incorporates the encoding layers of FastFormer into the XLM-Roberta framework. FastFormer generates contextual embeddings using additive attention mechanisms, allowing for extracting more effective contextual features. Experimental results show that the proposed approaches improve accuracy compared to the state-of-the-art by 1.5% and 0.9% in sentiment detection and 3.9% and 1.97% in emotion detection on the publicly available SentiMix code-mixed benchmark dataset.

1 INTRODUCTION

In today's digital world, social media platforms like X, Facebook, WhatsApp, and TikTok have become powerful tools for public discourse, shaping conversations around everything from politics to entertainment. Twitter alone reports around 187 million daily active users[1], many of whom contribute to a global exchange of ideas. In this rapidly evolving online landscape, India has emerged as a significant player, with millions of its citizens using these platforms to voice their opinions (Reddy and Muralidhar, 2021). However, the rich linguistic diversity in India poses a new challenge of frequently mixing languages, mainly Hindi and English, in their posts. This phenomenon, known as "code-mixing," has given rise to a new form of expression, Hinglish, that is becoming increasingly common on social media in India.

This blend of languages allows users to communicate more naturally in their multilingual contexts, and it poses a significant challenge for automated systems tasked with analyzing the sentiment behind these posts. Sentiment analysis —classifying a post as positive, negative, or neutral— is already a complex task, which becomes even more difficult in code-mixed languages like Hinglish. The nuances of emotion, the tone of a message, or even its intent can easily get lost in translation, leading to a misunderstanding of the underlying meaning.

Original message: "@RubikaLiyaquat Ek dum sahi kaha TV news ke jariye naam kaam liya paisa bana liya or ab boycott ka natak ... Joker"

English Translation: "@RubikaLiyaquat You said it absolutely right, you got name and fame through TV news, made money and now the drama of boycott...you joker"

The above example shows the complex nature of language mixing, where Hindi and English are seamlessly combined within a single post. Although

[1] https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/

Kumar, A., Pandey, A., Ahlawat, S. and Prasad, Y.
On Enhancing Code-Mixed Sentiment and Emotion Classification Using FNet and FastFormer.
DOI: 10.5220/0013173600003890

In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 3*, pages 670-678
ISBN: 978-989-758-737-5; ISSN: 2184-433X

a human reader with a background understanding of Hindi and English can easily interpret the sentiment and detect the underlying emotions, this task is significantly more challenging for machine learning models. The message mixes two languages and incorporates informal and idiomatic expressions with implicit meaning. Understanding the true sentiment behind a phrase like "boycott ka natak" (the drama of boycott) requires cultural knowledge, as well as an understanding of how sarcasm or frustration is being conveyed. For machine learning models, this blend of language, culture, and context creates a challenge because the difficulty is compounded by the need to process texts in which linguistic norms are not strictly followed and where abbreviations, emojis, and cultural references fluidly co-exist with standard language

Traditional sentiment classification methods, such as aspect and document level analysis, have shown effectiveness in single-language contexts (Xu et al., 2019; Yin and Chang, 2020; Tang et al., 2022), and recently pre-trained language models, such as BERT and RoBERTa, have achieved impressive results in sentiment analysis (Matthew et al., 2018; Radford and Narasimhan, 2018; Devlin et al., 2019) in multilingual contexts. However, these models often falter when applied to code-mixed texts due to their training on primarily monolingual datasets. Soumitra et al. (Ghosh et al., 2023) addressed this challenge by introducing a code-mixed dataset with sentiment and emotion labels and applying a transfer learning approach using XLM-Roberta (Conneau et al., 2020). Although this model achieved good accuracy, detecting subtle emotions in code-mixed texts remains challenging due to the nuanced nature of emotions like sarcasm or frustration, which demand models capable of learning intricate features.

Two methodologies were proposed to address these challenges: the XLMR-FNet and XLMR-FastFormer models. The first approach enhances the model's ability to capture complex patterns by integrating a Fourier Mixing Sublayer and an FNet transformer encoder layer (Lee-Thorp et al., 2022) into the XLM-RoBERTa architecture. The second approach improves attention calculations and input representations by incorporating FastFormer layers (Wu et al., 2021) with XLM-RoBERTa, enabling better context capture. Both approaches improve the accuracy and F1-score for sentiment and emotion classification on the publicly available SentiMix code-mixed benchmark dataset. Further details on the experiments & results are provided in section 5.

The key contributions of this work are as follows:

- To capture more intricate features such as frequency-domain information, the Fourier Mixing Sublayer, and Fourier Transform encoder are integrated with the cross-lingual embedding model XLM-Roberta.
- To leverage transfer learning, additive attention-based contextual embedding is integrated with the cross-lingual embedding model XLM-Roberta.
- Extensive experiments have been conducted on the SentiMix dataset to show the effectiveness of the proposed models against the SOTA models.

Furthermore, this paper is organized into five more sections: Literature Review, Preliminaries, Methodology, Experiments and Results, and Conclusion.

2 LITERATURE REVIEW

Increasing the presence of code-mixed languages such as Hinglish (Hindi-English) on social media platforms and analyzing the sentiments and emotions embedded in these texts presents unique challenges for natural language processing (NLP) systems. Sentiment analysis methods, which classify text into broad categories like positive, negative, or neutral, are often inadequate for handling such mixed-language input. The complexity increases when the task extends to emotion analysis, which requires models to identify specific emotional expressions, such as joy, anger, or frustration. Both sentiment and emotion analysis are crucial in NLP and have significant applications in various industries, including campaign monitoring (HaCohen-Kerner and Yaakov, 2019) and marketing (Sandoval et al., 2020). As code-mixed language becomes more widespread in digital communication, there is an urgent need for models that can effectively interpret these intricate linguistic patterns. The importance of accurate sentiment and emotion classification cannot be overstated. Traditional approaches to sentiment analysis, such as Support Vector Machines (SVM) and Naive Bayes, have been widely used. In (Agarwal, 2005), classification accuracy was improved using graph cut techniques on Wordnet graphs with synonyms. Over time, neural network-based methods were adopted to enhance performance (Socher et al., 2013; Patra et al., 2018), with studies like (Ghosh et al., 2017) exploring hate speech detection in code-mixed Hinglish texts. Other research has focused on sentiment classification in code-mixed texts using Long Short-Term Memory (LSTM) networks with sub-word-level representations (Joshi et al., 2016; Kazuma et al., 2017).

Further advancements included using transliteration techniques for classification (Wang et al., 2012;

Table 1: Sentiment class distribution across training, testing, and validation sets.

Split	Total	Positive	Neutral	Negative
Train	14,000	4,634	5,264	4,102
Test	3,000	1,000	1,100	900
Validation	3,000	982	1,128	890

Shekhar et al., 2020), as well as translation and multitask learning approaches (Mathur et al., 2018; Caruana, 1997). Joint learning approaches were investigated in (Kumar et al., 2019; Akhtar et al., 2019; Akhtar et al., 2022), showing that they could enhance model performance. Recently, transformer models have proven highly effective in NLP tasks. Research such as (Wadhawan and Aggarwal, 2021; Vijay et al., 2018; Ghosh et al., 2023) demonstrated that transformer models outperform earlier sentiment and emotion analysis methods, particularly in code-mixed data. The development of cross-lingual transformer models like XLM-RoBERTa (Conneau et al., 2020), trained on 104 languages, has significantly improved sentiment analysis tasks in multilingual environments, showing good performance in both monolingual and cross-lingual benchmarks. The research by Ghosh et al. (Ghosh et al., 2023) tackled challenges in emotion and sentiment analysis for code-mixed data using a multitask transformer-based framework. They created an emotion-annotated Hindi-English code-mixed dataset derived from SentiMix and fine-tuned the XLM-RoBERTa model for sentiment and emotion classification. Tmultitaskingitask approach improved efficiency and outperformed benchmarks, highlighting the advantages of using emotion classification to enhance sentiment accuracy. However, models like XLM-RoBERTa still face difficulties capturing subtle emotion variations due to informal syntax, language switching, and cultural nuances in code-mixed texts. To address these, XLMR-FNet (Lee-Thorp et al., 2022) and XLMR-Fastformer (Wu et al., 2021) were introduced to enhance performance in these contexts.

Table 2: Emotion label distribution across the dataset.

Emotion	Train	Test	Validation
Anger	2,095	680	415
Disgust	1,048	105	148
Fear	56	13	4
Joy	3,893	1,008	973
Sadness	856	122	307
Surprise	51	7	6
Others	6,001	1,065	1,048
Total	**14,000**	**3,000**	**3,000**

XLMR-FNet uses a Fourier Transform instead of traditional self-attention, improving its ability to capture complex features, while XLMR-Fastformer employs an additive attention mechanism for efficient global context processing. These innovations enable better handling of large datasets and more accurate classification of nuanced emotions like sarcasm. Details on these models and their architectures are in the Methodology section, with experimental results demonstrating their effectiveness in code-mixed sentiment and emotion analysis provided in the Results section.

3 PRELIMINARIES

This section presents the mathematical formulation of the problem, followed by a detailed description of the dataset utilized for model training and evaluation.

3.1 Problem Definition

Definition: Given a set of messages $X = \{x_1, x_2, \ldots, x_n\}$, where X represents messages in a code-mixed language (T), the task is to train a model M such that: $E, S = M(X)$, where E refers to emotion, and S refers to sentiment. Emotions E are categorized into {anger, disgust, fear, joy, sadness, surprise, others}, and sentiment S is classified into {positive, negative, neutral}. The goal of the model M is to minimize the cross-entropy loss function, defined as follows:

$$E(\theta) = \sum_{c=1}^{N} y_{o,c} \log(p_{o,c}) \tag{1}$$

where θ represents the parameters of the model, $y_{o,c}$ denotes the true class label for the o^{th} message x_o, and $p_{o,c}$ is the predicted probability of the o^{th} message belonging to class c. The aim is to minimize this loss, encouraging the model to assign higher probabilities to the correct classes, thus improving classification accuracy.

For multitask learning, separate losses are computed for sentiment and emotion classification: $E_{sent}(\theta)$ for sentiment and $E_{emo}(\theta)$ for emotion. The combined loss is computed as a weighted sum of these individual losses:

$$E(\theta) = \alpha_{sent} E_{sent}(\theta) + \alpha_{emo} E_{emo}(\theta) \tag{2}$$

Where α_{sent} and α_{emo} represent the weight coefficients for sentiment and emotion, respectively. Using grid search, it was determined that setting both α_{sent} and α_{emo} to 0.3 produced the optimal results.

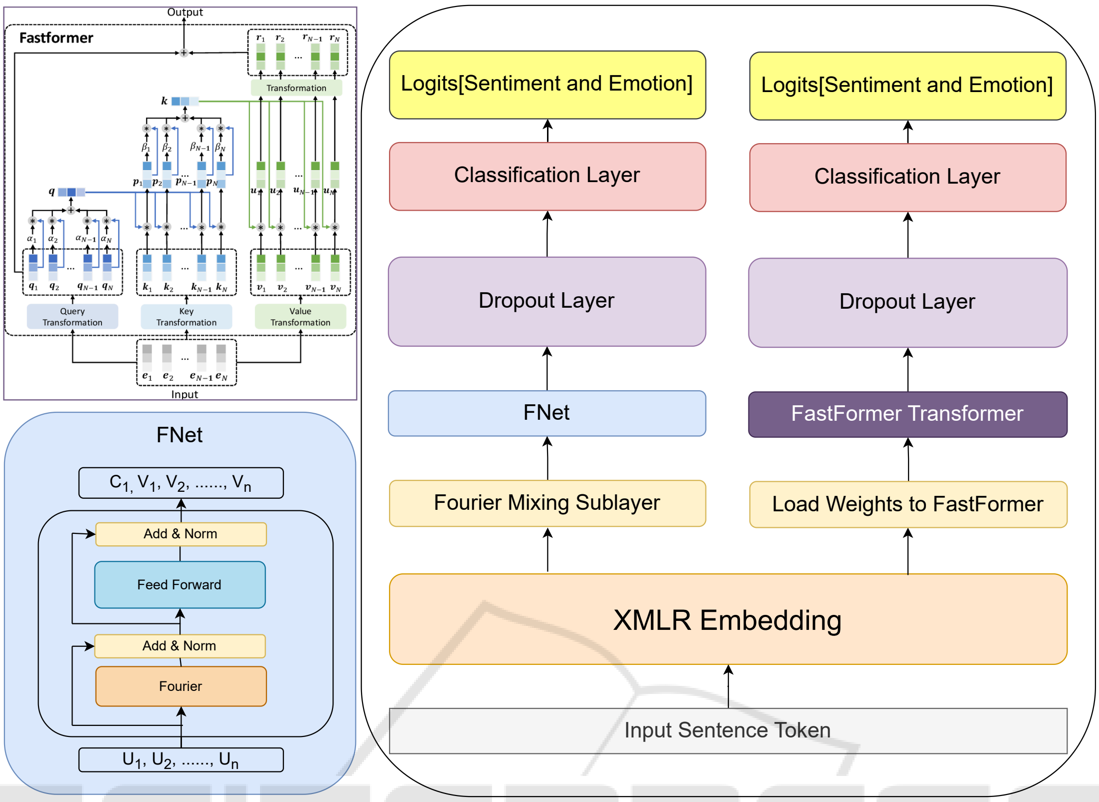

Figure 1: The above Framework shows both architectures FNet with XLMR and FastFormer with XLMR.

With the problem definition established, we now turn to the details of the dataset used to train and evaluate the proposed models, including label distributions and class balance.

3.2 Corpus Details

The dataset used in this study is from the SemEval 2020 shared task, specifically Task 9, which contains code-mixed Hindi-English (Hinglish) messages annotated with sentiment labels at the sentence level. The dataset is split into 14,000 instances for training, 3,000 for testing, and 3,000 for validation. Table 1 shows the class-wise sentiment distribution across these splits. The distribution indicates a slight bias toward the neutral class; overall, the dataset remains generally balanced.

The sentiment labels are categorized into three classes: positive, neutral, and negative. Positive messages express emotions such as happiness, gratitude, or appreciation, while negative messages convey criticism, harassment, or targeted attacks. Neutral messages generally contain information without a strong emotional undertone. For emotion classification, the dataset was further enriched by (Ghosh et al., 2023), who annotated the dataset with emotion labels derived from (Ekman et al., 1969) primary emotions: anger, disgust, fear, joy, sadness, and surprise. An additional category labelled "others" was introduced to capture non-emotive cases or instances that do not align with (Ekman et al., 1969) basic emotions. Table 2 presents the emotion label distribution across the training, testing, and validation splits.

4 METHODOLOGY

This section details the two models proposed in this work. The first model, XLMR-FNet, enhances XLM-RoBERTa by incorporating a custom FNet transformer encoder layer with a Fourier Mixing Sublayer to capture complex patterns in code-mixed text. The second model, XLMR-Fastformer, integrates XLM-RoBERTa with Fastformer layers, which use efficient attention mechanisms to improve contextual embeddings. Both models leverage transformer-based techniques and thorough data preprocessing to achieve robust multilabel text classification. This is mainly due to the Fourier Transform's unique ability to capture frequency-domain information. By applying a Fourier Transform, it transforms signals (e.g., timeseries or spatial signals) into their frequency domain representations. This allows the model to capture global patterns and periodicities in the data, which might be hard to detect in the raw time/space domain. The architectures of the XLMR-FNet and XLMR-Fastformer models are depicted in Figure 1 and further explained in sections 4.1 and 4.2.

4.1 XLMR with FNet

The XLMR-FNet architecture is designed to address sentiment and emotion analysis in a structured, five-step process. The input sentence is first tokenized using XLM-RoBERTa's tokenization mechanism, where tokens are assigned numerical representations through XLM-RoBERTa's embedding layer. The core innovation of this model lies in the second stage, where a Fourier Mixing Sublayer replaces the traditional self-attention mechanism found in standard transformers. This sublayer leverages the Discrete Fourier Transform (DFT) to decompose the input sequence into constituent frequency components, enabling the model to capture complex patterns in the data. The DFT for a sequence x_n (where n ranges from 0 to $N-1$) is defined as:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} nk}, \quad 0 \le k \le N-1$$

This equation represents the frequency domain representation for frequency k, where the original tokens (x_n) are transformed into frequency components. The Fourier Mixing Sublayer thus allows the model to analyze the sentence from multiple frequency perspectives, capturing subtle emotional nuances inherent in the code-mixed text. Following this, the model passes through the FNet layer, which is essential for learning complex relationships between the encoded tokens. A dropout layer is then applied to mitigate overfitting, ensuring that the model generalizes well to unseen data. The classification layer predicts both the sentiment and emotion of the input sentence, with the final loss calculated as a combination of cross-entropy errors for both sentiment and emotion (refer to equation 2). This loss guides the model to optimize both tasks simultaneously. The Fourier Transform's ability to capture hidden patterns makes this architecture particularly adept at handling the intricate language switching found in code-mixed texts. The effectiveness of this architecture in addressing the challenges of code-mixed sentiment and emotion analysis is demonstrated in the Results section.

4.2 XLMR with FastFormer

The XLMR-Fastformer model builds on the architecture of FastFormer, optimized for multilabel classification tasks. By integrating XLM-RoBERTa's contextual embeddings with the efficient attention mechanisms of FastFormer, this model enhances performance in sentiment and emotion classification tasks. The model begins by initializing tokenized sentences with XLM-RoBERTa embeddings, followed by FastFormer layers, which are initialized with pre-trained XLM-RoBERTa weights.

A key component of FastFormer is its additive attention mechanism, which reduces computational complexity while preserving the ability to capture global contextual information. This mechanism is efficient compared to the standard self-attention used in traditional transformers, making it particularly suitable for large-scale datasets. The process involves generating query (Q), key (K), and value (V) matrices from the input sequence X, defined as:

$$Q = \text{Linear}_Q(X), K = \text{Linear}_K(X), V = \text{Linear}_V(X)$$

Unlike standard transformers, FastFormer applies additive attention, which summarizes the query matrix into a global query vector via attention weights:

$$\alpha = \text{softmax}\left(\frac{QW_Q}{\sqrt{d}}\right), \quad Q_{\text{global}} = \sum_{i=1}^{N} \alpha_i Q_i,$$

Here, α is the attention weight, and d represents the input dimension. The global query vector Q_{global} interacts with the key matrix, forming an intermediate matrix P, which is further summarized into a global key vector:

$$K_{\text{global}} = \sum_{i=1}^{N} \beta_i P_i$$

Finally, the global key vector interacts with the value matrix, producing the final output, which is transformed by a linear layer to generate the interaction output. This output is then added back to the original input sequence to produce the final output of FastFormer layer. The overall FastFormer architecture consists of multiple such layers stacked together, using dropout for regularization.

FastFormer architecture is highly efficient due to its reduced computational complexity, making it scalable for large datasets. Its ability to capture global context is especially beneficial for code-mixed language tasks, where context-switching between languages is common. Similar to XLMR-FNet, the final

Table 3: This table shows Total Accuracy (Accuracy) and Weighted F1 scores (F1 Score) with different models. The highest scores are shown in bold. The − represents the unavailability of results. MTL and TL represent Multitask Learning and Transfer Learning, respectively.

.

Tasks	Sentiment		Emotion	
Methods	Accuracy	F1 Score	Accuracy	F1 Score
State of the art Baselines				
XLM	69.16	69.20	65.20	62.29
mBERT (George-Eduard et al., 2020)	68.66	69.15	–	–
XLM MTL (Malte et al., 2020)	70.47	70.48	65.20	62.29
TL-XLM (Ghosh et al., 2023)	71.30	71.61	66.03	64.47
Proposed Methods				
XLM-FastFormer	**72.20**	70.97	**68.00**	**65.83**
XLM-FNet	**72.80**	71.37	**69.93**	**68.01**

Table 4: Some cases where the previous state of the art with transfer learning failed and the correct results were highlighted.

Message	**Label**	**FNetXLMR**	**FastFormerXLMR**	**TLXLMR**
Original Text: @akki@iRajeev You are stalking me bro. Just shut up and nikal. Le. Faltu logo ki mention mein jagah nahi haii Translated Text: @akki@iRajeev You are stalking me, bro. Just shut up and get lost. Take this. There's no place for useless people in the mentions	Sentiment: **negative** Emotion: **disgust**	Sentiment: **negative** Emotion: anger	Sentiment: **negative**, Emotion: anger	Sentiment: neutral Emotion: anger
Original Text: I found this awesome recording of Aawaz deke hamein tum bulao on #Smule Translated Text: I found this awesome recording of 'Aawaz deke hamein tum pulao' on #Smule	Sentiment: **neutral** Emotion: **joy**	Sentiment: **neutral** Emotion: **joy**	Sentiment: **neutral**, Emotion: **joy**	Sentiment: positive Emotion: others
Original Text: @nirahua1 @msunilbishnoi Wah bhai Shandaar jabab Translated Text: @nirahua1 @msunilbishnoi Wow, brother, great response	Sentiment: **positive** Emotion: **joy**	Sentiment: **positive** Emotion: **joy**	Sentiment: **positive**, Emotion: **joy**	Sentiment: neutral Emotion: others

loss is computed by combining the cross-entropy errors for both sentiment and emotion classification, as detailed in the Results section. This model excels in tasks requiring the processing of extensive data while maintaining high accuracy. Further details on the performance improvements of XLMR-Fastformer can be found in the Results section, where we provide evidence supporting the efficiency and accuracy gains achieved by this architecture.

5 EXPERIMENTS AND RESULTS

This section presents the experimental setup and results, demonstrating the effectiveness of the proposed models compared to state-of-the-art techniques. The results are analyzed using performance metrics such as Accuracy and F1-score, showcasing the superior performance of the models in multilabel text classification for sentiment and emotion analysis in code-mixed data.

5.1 Experiment Setup

The training pipeline comprised extensive data preprocessing and tokenization steps, followed by model training using the AdamW optimizer in conjunction with CrossEntropyLoss. Key hyperparameters were carefully tuned to ensure optimal performance. The final configuration included the following settings: random seed of 42, GPU configuration with an A100 40GB, a pre-trained XLM-RoBERTa model, dropout rate of 0.5, batch size of 16, learning rate of 2×10^{-5}, a maximum sequence length of 128 tokens, 10 epochs (with early stopping after 2 epochs of no improvement), and loss weights of 0.3 for sentiment and emotion tasks.

5.2 Results

The experimental results, as shown in table 3, confirm the effectiveness of our proposed models, XLMR-FNet and XLMR-FastFormer. The first method enhances the model's ability to identify and process complex patterns by incorporating a Fourier Mixing Sublayer along with a new FNet transformer encoder layer into XLM-Roberta, effectively decompose these complex patterns into frequency components, enabling the model to detect the underlying linguistic shifts. This modification results in a significant improvement in the model's performance, with sentiment accuracy reaching 72.8% and emotion accuracy achieving 69.93%. These results represent an increase of 1.5% in sentiment accuracy and 3.9% in emotion accuracy compared to previously established benchmarks. Furthermore, the second method focuses on refining the additive attention mechanism to capture global context and provide better input representation, which is crucial for accurate sentiment and emotion classification. By integrating Fastformer layers into the XLM-Roberta architecture, this approach achieves a sentiment accuracy of 72.2% and an emotion accuracy of 68%. Although the improvements over the baseline are more modest—0.9% in sentiment accuracy and 1.97% in emotion accuracy—they still demonstrate the effectiveness of this approach in enhancing the model's performance.

Despite the strong overall performance, the confusion matrices in Figure 2 highlight areas where challenges persist. For instance, in XLMR-FNet, 253 instances of neutral sentiment were misclassified as negative. Similarly, XLMR-FastFormer also exhibited confusion between neutral and negative, with 262 neutral instances being misclassified as negative. This difficulty likely stems from the subtle overlap between neutral and negative sentiment, particularly in code-mixed data, where linguistic and contextual features may blur the boundaries between these categories. These results indicate that while our models reduce misclassifications compared to baselines, further improvements are needed in distinguishing between these closely related sentiments. On the contrary, in case of emotion classification, both models showed improved performance over the baselines. However, certain challenges remain in separating some emotions, particularly joy, and others. XLMR-FNet misclassified 153 instances of joy as others, and XLMR-FastFormer, though slightly better, still misclassified 128 instances. These results suggest that while our models capture most of the nuanced emotional variation present in code-mixed texts, further refinement is needed to fully differentiate between these overlapping emotional categories.

To further demonstrate the effectiveness of the proposed models, Table 4 provides specific inference examples where our models succeeded in cases where the baselines failed. For example, in one case, XLMR-FNet correctly identified a Message with negative sentiment and disgust emotion, whereas TL-XLM incorrectly classified it as neutral. This showcase how the Fourier Transform in XLMR-FNet captures hidden linguistic patterns that other models miss, particularly in informal and contextually rich texts. Another case illustrates how XLMR-FastFormer's attention mechanism captured subtle emotional features, correctly classifying joy with a neutral sentiment where baseline models misclassified it as positive. Furthermore, our models have shown significant improvements, so it's important to acknowledge certain trade-offs. The Fourier Transform and attention mechanisms, though effective, come with increased computational complexity, particularly in terms of time and resource consumption. Nevertheless, these trade-offs are justified by the models' enhanced ability to capture complex linguistic patterns and improve performance across both sentiment and emotion classification tasks. The improvements in accuracy, particularly in handling complex code-mixed texts, show the potential of our models to address the challenges of multilabel classification in sentiment and emotion analysis.

6 CONCLUSION

This paper introduces two enhanced transformer-based methodologies that integrate XLM-RoBERTa with Fourier Mixing and FastFormer layers, addressing the challenges of suboptimal performance due to insufficient feature extraction in code-mixed language processing. The proposed methods achieve significant advancements in sentiment and emotion classification accuracy for code-mixed texts, outperforming existing state-of-the-art methods.

However, this work has certain limitations. The experiments are limited to the SentiMix dataset with Hindi-English (Hinglish) code-mixed text, which restricts the generalizability of the findings. While the performance improvements are notable, the computational overhead introduced by the Fourier and FastFormer layers may hinder their practicality for large-scale real-world applications. Future work should

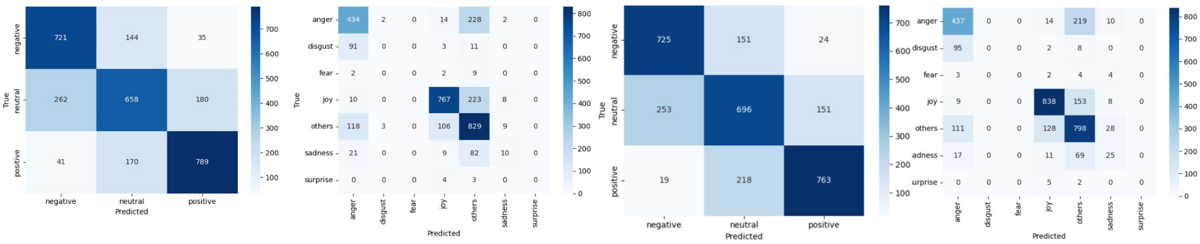

Figure 2: Confusion Matrices for Sentiment and Emotion detection of FastFormer (*left two matrices*) and FNet Model (*right two matrices*).

explore self-supervised pretraining, optimize attention mechanisms, and enhance efficiency for under-resourced code-mixed languages.

ACKNOWLEDGMENTS

This research is partially funded by IHUB NTIHAC FOUNDATION under project numbers IHUB-NTIHAC/2021/01/14 & IHUB-NTIHAC/2021/01/15.

REFERENCES

Agarwal, A. (2005). Sentiment analysis: A new approach for effective use of linguistic knowledge and exploiting similarities in a set of documents to be classified.

Akhtar, M. S., Chauhan, D., Ghosal, D., Poria, S., Ekbal, A., and Bhattacharyya, P. (2019). Multi-task learning for multi-modal emotion recognition and sentiment analysis.

Akhtar, M. S., Ghosal, D., Ekbal, A., Bhattacharyya, P., and Kurohashi, S. (2022). All-in-one: Emotion, sentiment and intensity prediction using a multi-task ensemble framework. *IEEE Transactions on Affective Computing*.

Caruana, R. (1997). Multitask learning. *Machine Learning*.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., and Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale.

Devlin, J., Ming-Wei, C., Kenton, L., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding.

Ekman, P., Sorenson, E. R., and Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*.

George-Eduard, Z., George-Alexandru, V., Dumitru-Clementin, C., Traian, R., and Costin-Gabriel, C. (2020). Upb at semeval-2020 task 9: Identifying sentiment in code-mixed social media texts using transformers and multi-task learning.

Ghosh, S., Ghosh, S., and Das, D. (2017). Sentiment identification in code-mixed social media text.

Ghosh, S., Priyankar, A., Ekbal, A., and Bhattacharyya, P. (2023). Multitasking of sentiment detection and emotion recognition in code-mixed hinglish data. *Knowledge-Based Systems*, 260:110182.

HaCohen-Kerner and Yaakov (2019). Text classification and sentiment analysis in social media for the marketing domain. In *International Journal of Research and Analytical Reviews (IJRAR)*.

Joshi, A., Prabhu, A., Shrivastava, M., and Varma, V. (2016). Towards sub-word level compositions for sentiment analysis of Hindi-English code mixed text.

Kazuma, H., Caiming, X., Yoshimasa, T., and Socher, R. (2017). A joint many-task model: Growing a neural network for multiple nlp tasks.

Kumar, A., Ekbal, A., Kawahra, D., and Kurohashi, S. (2019). Emotion helps sentiment: A multi-task model for sentiment and emotion analysis. In *2019 International Joint Conference on Neural Networks (IJCNN)*.

Lee-Thorp, J., Ainslie, J., Eckstein, I., and Ontanon, S. (2022). Fnet: Mixing tokens with fourier transforms.

Malte, A., Bhavsar, P., and Rathi, S. (2020). Team_Swift at SemEval-2020 task 9: Tiny data specialists through domain-specific pre-training on code-mixed data.

Mathur, P., Sawhney, R., Ayyar, M., and Shah, R. (2018). Did you offend me? classification of offensive tweets in Hinglish language.

Matthew, P., Neumann, M., Iyyer, M., Gardner, M., Christopher, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. In *Association for Computational Linguistics*.

Patra, B., Das, D., and Das, A. (2018). Sentiment analysis of code-mixed indian languages: An overview of sail code-mixed shared task icon 2017.

Radford, A. and Narasimhan, K. (2018). Improving language understanding by generative pre-training.

Reddy and Muralidhar (2021). Social media: Internet trends in india and growth of social media in the recent times. *International Journal of Business and Administration Research*.

Sandoval, A., Rodrigo, and ValleCruz, D. (2020). Sentiment analysis of facebook users reacting to political campaign posts. *Digit. Gov.: Res. Pract.*

Shekhar, S., Sharma, D., and Beg, S. M. (2020). An effective bi-lstm word embedding system for analysis and identification of language in code-mixed social media text in english and roman hindi.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., and Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank.

Tang, J., Xue, Y., Wang, Z., Hu, S., Gong, T., Chen, Y., Zhao, H., and Xiao, L. (2022). Bayesian estimation-based sentiment word embedding model for sentiment analysis. *CAAI Transactions on Intelligence Technology*.

Vijay, D., Bohra, A., Singh, V., Akhtar, S. S., and Shrivastava, M. (2018). Corpus creation and emotion prediction for Hindi-English code-mixed social media text.

Wadhawan, A. and Aggarwal, A. (2021). Towards emotion recognition in hindi-english code-mixed data: A transformer based approach.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., and Narayanan, S. (2012). A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle.

Wu, C., Wu, F., Qi, T., Huang, Y., and Xie, X. (2021). Fastformer: Additive attention can be all you need.

Xu, H., Liu, B., Shu, L., and Yu, P. (2019). BERT post-training for review reading comprehension and aspect-based sentiment analysis. In *Association for Computational Linguistics*.

Yin, D. and Chang, T. M. K.-W. (2020). SentiBERT: A transferable transformer-based architecture for compositional sentiment semantics. In *Association for Computational Linguistics*.