Hybrid Transformer–Recurrent Modelling for Sentiment Analysis in Low Resource Language

Sheezan Farooq
Department of Computer Science, Islamic University of Science and Technology, Jammu and Kashmir, India

Rumaan Bashir
Department of Computer Science, Islamic University of Science and Technology, Jammu and Kashmir, India

ABSTRACT
Sentiment analysis in low-resource languages like Kashmiri is underexplored due to the lack of annotated datasets and computational tools. This research proposes an effective hybrid deep learning model that combines the strengths of XLM-RoBERTa and BiLSTM networks for sentiment analysis of Kashmiri language text. The Kashmiri language, being low-resource and morphologically rich, poses significant challenges for natural language understanding tasks. A major contribution of this work is the creation of a manually annotated sentiment dataset tailored for the Kashmiri language, encompassing positive, negative, and neutral sentiment categories. This dataset serves as a foundational resource for training and evaluating sentiment classification models in this underrepresented language. The hybrid model combines XLM-RoBERTa transformer contextual embeddings with BiLSTM modeling sequences because of its ability to function in low-resource environments. Experimental results demonstrate that the hybrid model achieves state-of-the-art performance with a validation accuracy of 94.7% and an F1-score of 0.94 across all sentiment classes. Additionally, ROC analysis confirms high discriminative ability with an AUC of 0.99 for each class. Experimental findings demonstrate that integrating pre-trained transformers with recurrent models improves sentiment recognition abilities in Kashmiri by a significant degree.

Keywords
Kashmiri language, sentiment analysis, XLM-RoBERTa, BiLSTM, hybrid model, low-resource languages

1. INTRODUCTION
The method known as sentiment analysis or opinion mining functions as a necessary tool for understanding subjective information contained in text-based data[1]. The identification of opinions from text has transformed into a vital analytic solution which assists organizations in multiple fields including business intelligence and social media monitoring[2]. User-generated content expansion drives market demand for sentiment analysis solutions which need to function across multiple languages[3]. While sentiment analysis has significantly advanced for high-resource languages such as English and Chinese, NLP research continues to underrepresent low-resource languages, including Kashmiri[4]. The Kashmiri language is spoken by over 7 million people, primarily in the Kashmir Valley. However, the development of NLP tools for Kashmiri is hampered by a lack of annotated corpora, language-specific models, and linguistic resources[5]. A solution to combat these challenges requires a sentiment classification framework that is specifically designed for Kashmiri language processing. The approach leverages XLM-RoBERTa, a state-of-the-art multilingual transformer model pre-trained on a diverse corpus, and combines it with a Bi-directional Long Short-Term Memory (BiLSTM) network. This hybrid architecture enables the model to capture rich contextual relationships and sequential patterns in the text, making it particularly effective for nuanced sentiment classification. This work represents one of the first comprehensive studies focused on deep learning-based sentiment analysis for Kashmiri. By developing and evaluating a hybrid XLM-RoBERTa + BiLSTM model, the study aims is to contribute the growing body of research on NLP for under-resourced languages and provide a foundation for future advancements in this area.

2. RELATED WORK
Research on sentiment analysis for Indian languages such as Hindi, Tamil, and Bengali has gained traction with the help of different machine learning, deep learning and multilingual transformers like mBERT and XLM-R[6], [7], [8]. Early research efforts depended on Naive Bayes along with Support Vector Machines (SVM) [9] and Decision Trees through conjunction with TF-IDF or Bag-of-Words features[10]. Deep learning models surpassed previous methods because they failed to identify semantic meaning in text. Recent analysis utilizes transfer learning and multilingual models[11], [12]. For instance, mBERT has been used for sentiment classification in Hindi, Bangla and Tamil, achieving notable improvements over traditional methods[13], [14]. XLM-RoBERTa has shown superior performance in multilingual NLP tasks due to its larger training corpus and deeper architecture[15]. Hybrid models that combine transformer embeddings with recurrent layers like LSTM or GRU have also gained popularity[16], [17]. Hybrid models combine transformer understanding with RNN sequential processing to deliver better results for sentiment analysis and emotion detection alongside intent classification[13]. However, no significant effort has been reported for Kashmiri language sentiment analysis using such advanced hybrid architectures. This work fills research gap by introducing a transformer-RNN hybrid model specifically tailored for the Kashmiri language.

3. METHODOLOGY
3.1 Dataset Collection and Preprocessing
In order to create the Kashmiri Sentiment Analysis Dataset, 19,998 manually chosen English sentences were collected from benchmark datasets[18], online reviews, news articles, and social media. These sentences were then translated to Kashmiri using Microsoft Bing Translator, and an expert review was conducted to ensure accuracy. Three sentiment classes Positive (6,666), Negative (6,666), and Neutral (6,666) were carefully balanced throughout the dataset. Three columns make up the dataset: Serial No., Kashmiri Sentence, and Sentiment. UTF-8 encoding was used to export the final dataset to a CSV file so that it would function correctly with text processing applications. Table 1 provides a few examples of dataset. This dataset offers efficient deep learning-based sentiment classification for Kashmiri text. Initially, whitespaces were

manually removed and missing (NaN) values were also handled in the dataset. Following this, preprocessing was conducted to meet the requirements of the hybrid deep learning model combining XLM-Roberta and BiLSTM. The input text was first tokenized using the pre-trained XLM-Roberta tokenizer, converting it into input IDs and attention masks. These embeddings, generated by the XLM-Roberta model, were then passed to a bidirectional LSTM layer to capture sequential dependencies. The final hidden states were further processed through a fully connected layer to perform the classification.

Table 1: Samples of Dataset

Sen_Number	English _Sentence	Kashmiri _Sentence	Sentiment
9994	I absolutely detest this meal; it's bad	بہٕ چُھس اتھ کھینس بالکل نفرت کران۔ یہِ چُھ خراب۔	NEGATIVE
9995	I can't believe how pathetic this feature is; ...	مےٚ ہیٚکو نہٕ یقین کُرتھ زِ یہٕ خصوصیت کوتاہ ا...	NEGATIVE
9996	I absolutely complain about this plan; it's ho...	بہٕ چُھس اتھ منصوبس متعلق بالکل شکایت۔ یہٕ چُھ...	NEGATIVE
9997	This is the most incredible plan I've ever enc...	یہِ چھُ سارِوی کھوتہٕ زِیٛادٕ ناقابل یقین منصو...	POSITIVE
9998	The river looks intact as expected.	دُریاو چُھ توقعہٕ مُطأبِق برقرار باسان۔	NEUTRAL

3.2 Model Architecture

The proposed model architecture comprises several components. First, input tokens are processed by XLM-RoBERTa, a transformer-based encoder that generates deep contextual embeddings. The embedding block includes word-piece embeddings, position embeddings, and token type embeddings followed by layer normalization and dropout. These embeddings are then passed through 12 encoder layers of the XLM-RoBERTa backbone. The output representation corresponding to the [CLS] token is extracted and further refined using the model's internal pooler (a fully connected layer followed by a Tanh activation). To capture sequential information, the [CLS] token output is passed through a single-layer BiLSTM with a hidden size of 512. The BiLSTM layer effectively learns the forward and backward dependencies in the sentence. The final BiLSTM output is regularized using dropout and then fed into a fully connected linear layer that projects the 512-dimensional representation down to 3 output units (representing the three sentiment classes). A softmax activation is applied during training to compute class probabilities. The complete model has approximately 280 million trainable parameters, with most parameters belonging to the XLM-RoBERTa transformer. The architecture is optimized for GPU computation, achieving efficient memory usage with a total estimated size of 1228 MB during training.

3.3 Training Details

The hybrid model underwent a specially designed training process to achieve strong learning capabilities together with generalized performance. CrossEntropy Loss function is utilized, a standard choice for multi-class classification problems. To optimize the model parameters, the AdamW optimizer was employed, known for its adaptive learning rate and improved weight decay handling. To prevent the model from plateauing during training, the study integrated the ReduceLROnPlateau learning rate scheduler, which reduces the learning rate when a metric has stopped improving. The model was trained over 12 epochs with an initial learning rate of 2e-5. A dropout layer with a probability of 0.3 was applied after the BiLSTM layer to reduce overfitting and enhance generalization. The BiLSTM component was configured with a hidden size of 512, allowing it to effectively capture bidirectional dependencies in the input text. These hyperparameters were selected through extensive experimentation and tuning, resulting in optimal convergence and performance on the validation dataset.

4. EXPERIMENTAL RESULTS

4.1 Training and Validation Curves

The model's performance over 12 training epochs is illustrated in Fig. 1a and Fig. 1b. Fig. 1a presents the training and validation loss curves, while Fig. 1b depicts the corresponding training and validation accuracy curves. Initially, the training loss showed a sharp decline, indicating rapid learning. However, the validation loss exhibited some fluctuations particularly a spike around epoch 3 and a modest increase between epochs 6 to 8 before stabilizing in the later epochs. This behaviour suggests the model initially overfits slightly but later achieves more stable generalization as training progresses. On the other hand, training accuracy consistently improved across epochs, reaching nearly 97% by epoch 12. Validation accuracy followed a similar trend with minor dips, ultimately peaking above 94.7%. The narrowing gap between training and validation accuracy in the final epochs confirms that the model maintains strong generalization capability without overfitting. These curves collectively demonstrate the effectiveness of the training strategy and the robustness of the XLM-RoBERTa + BiLSTM architecture.

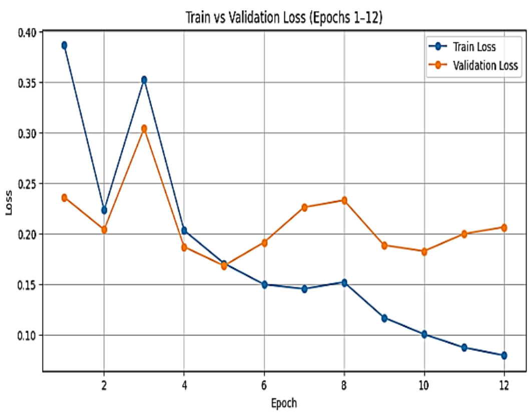

Fig 1a: Loss Curves of Hybrid Model

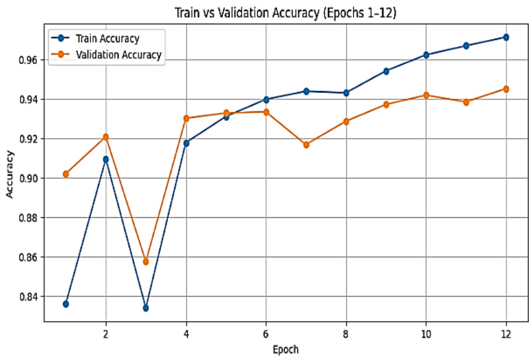

Fig 1: Accuracy Curves of Hybrid Model

4.2 Error Rate Analysis

The error rate trend over 15 epochs, shown in Fig 2, provides further insight into the model's learning behaviour. The training error decreased steadily from approximately 13.9% in the first epoch to below 2% by the final epoch. This sharp decline highlights the model's ability to fit the training data effectively. The validation error rate started slightly lower than the training error and remained relatively stable, consistently staying between 5% and 7%. While there were minor fluctuations, the error curve did not exhibit any signs of overfitting or instability. The consistent gap between training and validation error rates indicates that the model generalized well to unseen data.Overall, the error rate analysis reaffirms the stability and robustness of the proposed XLM-RoBERTa + BiLSTM model throughout the training process.

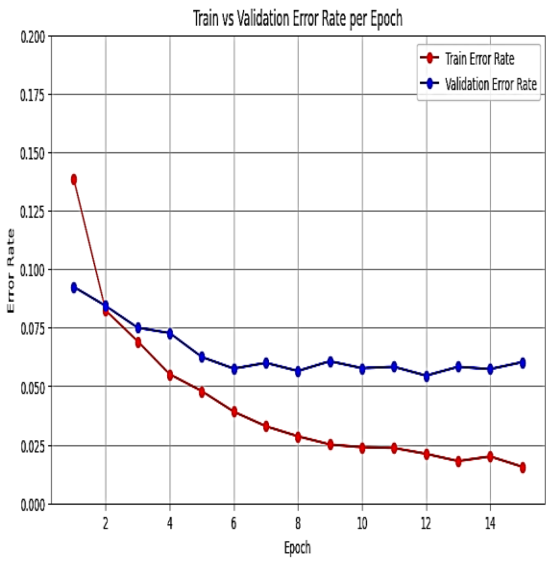

Fig 2: Training and Validation Error Rate of Hybrid Model

4.3 Classification Report

The performance of the proposed model is further substantiated by the classification report, which provides a detailed evaluation of precision, recall, and F1-score across the three sentiment classes Negative, Neutral, and Positive. As shown in table 2 below, the model achieves balanced and high scores for all metrics across each category:

i. Negative Class: Precision and recall both stand at 0.94, indicating the model's ability to correctly identify negative sentiments with minimal false positives and false negatives.

ii. Neutral Class: The model shows excellence when processing neutral text by achieving precision levels of 0.93 combined with recall levels of 0.94 which demonstrate its skill at detecting delicate sentiment indicators.

iii. Positive Class: A precision of 0.94 and a recall of 0.93 indicate that the model can effectively detect positive sentiments.

The overall accuracy of the model on the validation set is 94.7%, and the macro-averaged F1-score also stands at 0.94, confirming that the model performs consistently across all sentiment classes.

Table 2: Precision, Recall, F1-Score and Accuracy of Hybrid Model

Class	Precision	Recall	F1-Score	Support
Negative	0.94	0.94	0.94	1333
Neutral	0.93	0.94	0.94	1333
Positive	0.94	0.93	0.94	1334
Overall Accuracy	94.7%			

4.4 ROC Curve

The ROC curve in Fig 3 highlights the model's ability to distinguish between sentiment classes with high confidence. The Area Under the Curve (AUC) values for all three sentiment classes Negative, Neutral, and Positive are consistently 0.99, indicating near-perfect classification performance. The ROC curves for each class rise steeply toward the top-left corner of the plot, showing that the model achieves a high true positive rate with a low false positive rate. This suggests strong sensitivity and specificity across all classes. The overlap in performance further confirms that the model is not biased toward any particular sentiment class and handles class balance effectively. Overall, the ROC analysis reinforces the robustness and generalization capability of the XLM-RoBERTa + BiLSTM hybrid model for multiclass sentiment classification in a low-resource language setting.

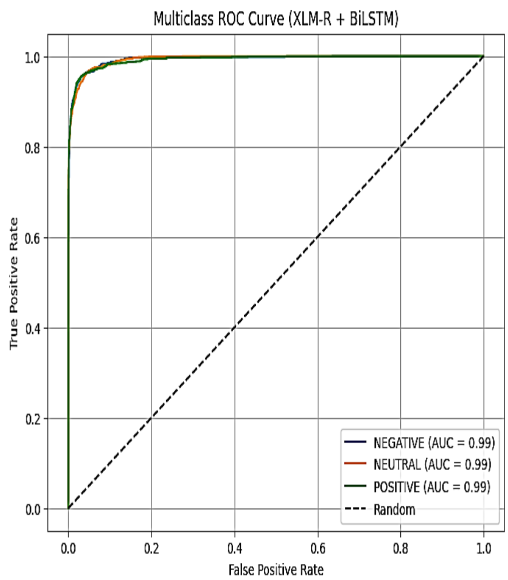

Fig 3: Multiclass ROC Curve of Hybrid Model

4.5 PERFORMANCE EVALUATION OF BASELINE MODELS

Table 3 presents the comparative performance of classical machine learning and deep learning models for Kashmiri sentiment classification. Among the evaluated models, Support Vector Machine (SVM) achieved the highest accuracy of 93.48% along with the best Macro-F1 score of 0.9343, indicating strong generalization capability across all sentiment classes. The LSTM model also demonstrated competitive performance with an accuracy of 92.84%, highlighting the effectiveness of sequential modeling in capturing contextual dependencies in Kashmiri text. Naïve Bayes (NB) achieved an accuracy of 92.27%, showing that traditional probabilistic methods remain reasonably effective for balanced datasets. Interestingly, mBERT, despite being a multilingual transformer model, achieved comparatively lower accuracy (92.45%). This may be attributed to limited Kashmiri-specific representation in the pre-training corpus, which can impact contextual understanding in low-resource settings. The results indicate that while transformer-based models offer contextual advantages, carefully tuned classical and recurrent architectures can perform competitively in low-resource language environments.

Table 3: Performance Comparison of Baseline Models

Model	Accuracy	Macro-Precision	Macro-Recall	Macro-F1
NB	0.9227	0.9239	0.929	0.9292
SVM	0.93448	0.93454	0.934	0.9343
LSTM	0.9284	0.9330	0.930	0.9300
mBERT	0.9245	0.9340	0.934	0.9300

5. CONCLUSION

This study presented a hybrid deep learning framework that integrates XLM-RoBERTa's contextual embedding capabilities with BiLSTM-based sequential modeling for sentiment classification of Kashmiri text. One of the primary contributions of this work is the development of a manually curated and balanced sentiment dataset for the Kashmiri language, addressing a significant resource gap in low-resource NLP research. A comprehensive comparative evaluation was conducted against classical machine learning models (Naïve Bayes, SVM) and deep learning architectures (LSTM and mBERT). Experimental findings demonstrate that while SVM and LSTM achieved competitive performance, the proposed hybrid XLM-RoBERTa + BiLSTM model outperformed all baseline approaches, achieving a validation accuracy of 94.7% and a Macro-F1 score of 0.94. The ROC analysis further confirmed strong discriminative capability with an AUC of 0.99 across all sentiment classes. The results indicate that combining multilingual transformer representations with sequential recurrent modeling enhances contextual understanding and sentiment discrimination in morphologically rich, low-resource languages such as Kashmiri. Extensive evaluation through training curves, error rate analysis, and multiclass ROC assessment confirms that the proposed model achieves strong generalization without overfitting. Overall, this research advances the development of robust NLP frameworks for underrepresented languages and provides a scalable foundation for future sentiment analysis and language technology applications in Kashmiri.

6. FUTURE WORK

In future studies, aim is to enhance this framework by introducing ensemble architectures that incorporate additional multilingual models such as IndicBERT and DistilBERT. The project will study domain-specific fine-tuning strategies alongside attention mechanisms and self-supervised learning approaches to enhance the sentiment classification outcomes. Additionally, expanding the dataset and incorporating dialectal variations of Kashmiri could improve linguistic coverage and increase the model's applicability in real-world settings.

Declarations

- **Conflict of Interest:** The authors of this manuscript declare there is no conflict of interest.
- **Funding**: The author states that this project has no associated funding.
- **Informed Consent:** Not applicable.
- **Data Availability:** on request to the corresponding author.
- **Authors' contributions**

  S.F: Conceptualization, Methodology, Formal Analysis, Data Creation, Writing Original draft, Funding acquisition. **R.B:** Conceptualization, writing review and editing, Supervision, Funding acquisition. All authors have read and approved the article for publication.

7. REFERENCES

[1] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," Electronics (Switzerland), vol. 9, no. 3, Mar. 2020, doi: 10.3390/electronics9030483.

[2] K. Rakshitha, R. H M, M. Pavithra, A. H D, and M. Hegde, "Sentimental analysis of Indian regional languages on social media," Global Transitions Proceedings, vol. 2, no. 2, pp. 414–420, Nov. 2021, doi: 10.1016/j.gltp.2021.08.039.

[3] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/j.asej.2014.04.011.

[4] M. R. R. Rana, A. Nawaz, and J. Iqbal, "A survey on sentiment classification algorithms, challenges and applications," Acta Universitatis Sapientiae, Informatica, vol. 10, no. 1, pp. 58–72, Aug. 2018, doi: 10.2478/ausi-2018-0004.

[5] N. A. Lone, K. J. Giri, and R. Bashir, "Natural Language Processing Resources for the Kashmiri Language," Indian J Sci Technol, vol. 15, no. 43, pp. 2275–2281, Nov. 2022, doi: 10.17485/IJST/v15i43.1964.

[6] S. Anbukkarasi and S. Varadhaganapathy, "Analyzing Sentiment in Tamil Tweets using Deep Neural Network," in Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020, Institute of Electrical and Electronics Engineers Inc., Mar. 2020, pp. 449–453. doi: 10.1109/ICCMC48092.2020.ICCMC-00084.

[7] M. R. H. Khan, U. S. Afroz, A. K. M. Masum, S. Abujar, and S. A. Hossain, "Sentiment Analysis from Bengali Depression Dataset using Machine Learning," in 2020

11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020, Institute of Electrical and Electronics Engineers Inc., Jul. 2020. doi: 10.1109/ICCCNT49239.2020.9225511.

[8] K. Shrivastava and S. Kumar, "A sentiment analysis system for the hindi language by integrating gated recurrent unit with genetic algorithm," International Arab Journal of Information Technology, vol. 17, no. 6, pp. 954–964, Nov. 2020, doi: 10.34028/iajit/17/6/14.

[9] M. Ahmad, S. Aftab, and I. Ali, "Sentiment Analysis of Tweets using SVM," Int J Comput Appl, vol. 177, no. 5, pp. 25–29, Nov. 2017, doi: 10.5120/ijca2017915758.

[10] N. Dhamayanthi and B. Lavanya, "The Role Of Naïve Bayes, SVM, And Decision Trees In Sentiment Analysis," Theory and Practice, vol. 2024, no. 4, pp. 6377–6381, 2024, doi: 10.53555/kuey.v30i4.2392.

[11] J. Y. Le Chan, K. T. Bea, S. M. H. Leow, S. W. Phoong, and W. K. Cheng, "State of the art: a review of sentiment analysis based on sequential transfer learning," Artif Intell Rev, vol. 56, no. 1, pp. 749–780, Jan. 2023, doi: 10.1007/s10462-022-10183-8.

[12] L. Khan, A. Amjad, N. Ashraf, and H. T. Chang, "Multi-class sentiment analysis of urdu text using multilingual BERT," Sci Rep, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-09381-9.

[13] A. S. Talaat, "Sentiment analysis classification system using hybrid BERT models," J Big Data, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40537-023-00781-w.

[14] Md. S. Islam and K. M. Alam, "Sentiment analysis of Bangla language using a new comprehensive dataset BangDSA and the novel feature metric skipBangla-BERT," Natural Language Processing Journal, vol. 7, p. 100069, Jun. 2024, doi: 10.1016/j.nlp.2024.100069.

[15] R. Cheruku, K. Hussain, I. Kavati, A. M. Reddy, and K. S. Reddy, "Sentiment classification with modified RoBERTa and recurrent neural networks," Multimed Tools Appl, vol. 83, no. 10, pp. 29399–29417, Mar. 2024, doi: 10.1007/s11042-023-16833-5.

[16] M. U. Salur and I. Aydin, "A Novel Hybrid Deep Learning Model for Sentiment Classification," IEEE Access, vol. 8, pp. 58080–58093, 2020, doi: 10.1109/ACCESS.2020.2982538.

[17] S. Zhou, Q. Chen, X. Wang, and X. Li, "Hybrid Deep Belief Networks for Semi-supervised Sentiment Classification."

[18] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification," Association for Computational Linguistics, 2007. [Online]. Available: http://ida.