Urdu Text Summarization Using Deep Learning

Hussain Umer Farooqui
School of Science and Engineering
Habib University
Karachi, Pakistan
hf08040@st.habib.edu.pk

Jazib Waqas
School of Science and Engineering
Habib University
Karachi, Pakistan
jw08048@st.habib.edu.pk

Muhammad Shayan Wasif
School of Science and Engineering
Habib University
Karachi, Pakistan
mw08071@st.habib.edu.pk

Ibrahim Rana
School of Science and Engineering
Habib University
Karachi, Pakistan
ir08043@st.habib.edu.pk

Dr Abdul Samad
School of Science and Engineering
Habib University
Karachi, Pakistan
abdul.samad@sse.habib.edu.pk

Sandesh Kumar
School of Science and Engineering
Habib University
Karachi, Pakistan
sandesh.kumar@sse.habib.edu.pk

***Abstract*—Abstractive text summarization generates concise summaries by understanding the meaning of the source text, rather than extracting sentences directly. Urdu, a low-resource language with complex linguistic features, presents unique challenges for this task. This paper explores the use of deep learning models to improve the quality of Urdu text summarization. We apply these models to a dataset of Urdu news articles, aiming to create accurate, coherent summaries while addressing the specific linguistic difficulties of Urdu.**

***Index Terms*—abstractive summarization, Urdu articles, deep learning, natural language processing**

I. Introduction

In today's digital world, the vast growth of content, particularly in news media, makes it essential to develop efficient text processing tools to help users digest large amounts of information quickly. While significant progress has been made in languages like English, Urdu remains underexplored in Natural Language Processing (NLP), despite being one of the most widely spoken languages in South Asia. Its rich linguistic heritage, influenced by Persian, Arabic, and Hindi, has not been matched by advanced NLP research, particularly in tasks like abstractive summarization.

Most existing Urdu summarization methods are extractive, selecting key sentences without fully capturing the essence of the text, often leading to fragmented results. In contrast, abstractive summarization generates a more coherent and human-like summary by paraphrasing and rephrasing the content. However, limited tools and resources exist for Urdu.

Our research aims to close this gap by developing a deep learning model for high-quality abstractive summarization of large Urdu texts. By leveraging transformer-based architectures like mBART, we seek to create a model that produces coherent, contextually accurate summaries, preserving the meaning of the original content. This work addresses a critical need in Urdu NLP, particularly in news media, improving the efficiency of information consumption for users.

Ultimately, this research contributes to broader efforts to develop advanced NLP tools for underrepresented languages, ensuring linguistic diversity in technological advancements.

II. Research Question and Problem Statement

The research question guiding our study is: "How can abstractive summarization of large Urdu texts be effectively achieved using deep learning models, specifically in low-resource language settings?" This inquiry addresses a pressing need for tools that can summarize lengthy Urdu texts, such as newspaper articles, into concise, fluent, and coherent summaries that capture the essential points of the original content.

The exponential growth of online Urdu content, particularly on news platforms, has created an overwhelming volume of information that users find challenging to process efficiently. Abstractive summarization offers a solution by generating paraphrased and restructured summaries, allowing users to grasp the key points of a text without reading its entirety. However, despite this need, Urdu remains significantly underrepresented in natural language processing (NLP). Current summarization methods for Urdu are limited, relying primarily on extractive techniques or classical machine learning models like Naïve Bayes and Support Vector Machines, which fail to capture the semantic richness required for abstractive summarization.

Generic multilingual models, such as mBART, while capable of processing low-resource languages like Urdu, are not without flaws. They often suffer from hallucination, generating irrelevant or incorrect content due to their lack of optimization for Urdu-specific tasks. Moreover, the scarcity of high-quality datasets for Urdu abstractive summarization presents a significant barrier to progress. Prior research has either lacked sufficient data or relied on inadequate methods, such as extractive summarization or generating overly simplistic summaries.

To overcome these challenges, our research leverages transformer-based architectures, particularly mBART, fine-tuned to handle the complexities of Urdu. We address the dataset scarcity problem by incorporating a large dataset of 64,000 Urdu news articles and summaries, complemented by a secondary dataset of 875 manually validated entries. This dual dataset approach ensures a diverse and high-quality training corpus. Additionally, our methodology prioritizes enhancing the summarization quality by focusing on rephrasing and restructuring rather than mere text reduction. This tailored approach minimizes the generic pitfalls faced by multilingual models and optimizes them for accurate and contextually relevant Urdu summarization.

By tackling these challenges, this research represents a significant advancement in Urdu NLP, addressing the limitations of prior models and paving the way for improved summarization tools in low-resource languages.

III. Literature Review

This literature review will look at various research papers that have contributed to the field of Urdu abstractive text summarization with particular focus on the deep learning models used, the data sets used and the evaluation measures adopted. It will point out the methods and results as well as the comparison of the approaches to be discussed, giving an idea of the further development of this field and the directions that can be further investigated.

Raza et al. [1] presented a method for abstractive summarization for Urdu with the help of transformer-based architectures. This study employed a news article dataset, although the size of the dataset and its structure were not clearly described. The authors mentioned that the data were enough to train the developed model in this study. For summarization, the manually created summaries were used as the ground truth. The entire data was divided into two sets, with 70% of data for training and 30% of data for testing. The method used included an encoder-decoder model, which is widely used in text summarization. Here, the encoder employed LSTM layers to encode the input text; the decoder produced the summary using an attention mechanism. Preprocessing included norming, stemming, tokenization and elimination of stop words. According to the ROUGE metrics, the proposed model obtained an F1 score of 0.43 for ROUGE-1, 0.25 for ROUGE-2, and 0.23 for ROUGE-L. These results showed that the model can produce good and useful summaries, which was the purpose of the model. The authors concluded that their proposed encoder-decoder framework along with the attention mechanism was promising for the abstractive summarization of low resource languages like Urdu.

In another study, Raza et al. [2] have also done work on abstractive summarization by using both extractive and abstractive methods. Their dataset was collected from leading Urdu newspapers like Express, BBC Urdu, Nawa-E-Waqt, Dawn, and Daily Jang and contains 50 articles of different topics including health, sports, politics, etc. The articles ranged from 400 to 1600 words, while summaries were between 33% and 80% of the article length. The authors suggested a combined solution based on the extractive techniques, including Sentence Weight Algorithm, TF-IDF, and Word Frequency Algorithm, as well as the abstractive BERT model. While creating extractive summaries, it was possible to select important sentences which accounted for 30-40% of the input text. For the abstractive approach, BERT, a pre-trained model was used in an encoder-decoder setup and the LSTM layers were used for the encoding and decoding. The model output was assessed using ROUGE indices even though the paper did not report the scores obtained. The authors pointed out that abstractive summaries were shorter and semantically more significant than the extractive ones, and the future research can be oriented on the improvement of the model to produce more concise summaries while preserving their logical structure.

Shafiq et al. [3] focused on abstractive text summarization for low-resource languages, particularly Urdu, using a dataset called the Urdu 1 Million News Dataset, which consisted of more than 1 million news articles categorized under sports, science and technology, business and economics, and entertainment. The dataset was split with 70% used for training and 30% for testing. The authors implemented a hybrid summarization approach that combined extractive techniques, such as the Sentence Weight Algorithm and TF-IDF, with abstractive techniques using a Seq2Seq model based on LSTM. The model included three encoder layers: one for the input sequence, one for keywords, and one for named entities. The encoder used a Bi-directional LSTM to capture both forward and backward context, while the decoder used a global attention mechanism to generate abstractive summaries. The model was evaluated using ROUGE metrics, with ROUGE-1 precision at 79%, recall at 30%, and an F1 score of 43%. The authors compared their results with existing Support Vector Machine (SVM) and Logistic Regression models, demonstrating that their deep learning model outperformed traditional methods. They also applied their model to Persian text summarization, where it outperformed traditional methods such as BERT and GPT.

Awais and Nawab [4] addressed the lack of large datasets by creating the UATS-23 Corpus, which included 2,067,784 Urdu news articles categorized into topics such as sports, entertainment, business, and science and technology. Each article contained a headline, treated as a summary, and the detailed story as the source text. The average number of words in the source text was 205.6, with an average of 9.39 words in the summaries. The maximum number of tokens in the source text was 3,000 words, and the minimum was 20 words. The authors employed multiple deep learning models, including LSTM, BiLSTM, GRU, and Bi-GRU, alongside transformer-based models like BART and GPT-3.5. Attention mechanisms were incorporated to allow the models to focus on relevant sections of the input. The GRU with attention model performed the best, achieving ROUGE-1 at 46.7%, ROUGE-2 at 24.1%, and ROUGEL at 48.7%. Transformer models such as BART and GPT-3.5 performed less effectively, highlighting challenges with n-gram overlap for Urdu summarization. The paper concluded that the UATS-23 corpus would serve as a valuable resource for future research on Urdu text summarization.

In another paper, Raza and Shahzad [5] presented an end-to-end solution for Urdu abstractive text summarization, introducing a dataset of 19,615 documents with corresponding summaries. The dataset was constructed by translating English documents and summaries into Urdu using Google Translate, followed by manual corrections to ensure accuracy. The model used a transformer-based architecture, which included RoBERTa embeddings to enhance contextual understanding of the text. The authors also introduced the Context-Aware RoBERTa Score (CA-RoBERTa Score), which combines cosine similarity and a disconnection rate to measure the coherence between sentences in the generated summary. The model was evaluated using both ROUGE and CA-RoBERTa metrics, achieving a ROUGE-1 score of 25.18%, a ROUGE-2 score of 12.14%, and a ROUGE-L score of 21.50%. The CA-RoBERTa score for the Urdu dataset was 20.61, reflecting strong sentence coherence in the summaries. The authors noted that the transformer model outperformed traditional approaches such as RNNs and LSTMs, but further improvements could be achieved by incorporating more advanced pre-trained models and fine-tuning techniques.

To summarize, the reviewed papers highlight the importance of combining deep learning models such as LSTM, GRU, and transformer-based architectures for generating high-quality summaries in Urdu. The scarcity of large, annotated datasets and pre-trained models continues to be a challenge, though efforts like the UATS-23 Corpus and novel metrics such as CA-RoBERTa provide a foundation for future work. Improvements in sentence coherence and evaluation metrics will be key to advancing Urdu abstractive summarization in low-resource settings.

Study	Models	Dataset	Results
[1]	Transformer-based (LSTM with Attention)	News articles dataset (size not mentioned)	ROUGE-1: 0.43 ROUGE-2: 0.25 ROUGE-L: 0.23
[2]	BERT	50 articles from various publications	Rogue score not specified.
[3]	(Seq2Seq with Bi-LSTM)	Urdu 1 Million News Dataset, Summaries extracted.	ROUGE-1: 43% ROUGE-2: 25% ROUGE-L: 23%
[4]	GRU with Attention	UATS-23 Corpus (2 million articles), headlines as summaries	ROUGE-1: 46.7%, ROUGE-2: 24.1%, ROUGE-L: 48.7%
[5]	Transformer-based (RoBERTa embeddings)	19,615 documents, translated and manually corrected	ROUGE-1: 25.18% ROUGE-2: 12.14%

Fig. 1. Summary of the literature review.

IV. Dataset

A. Data Acquisition:

The data acquisition process was crucial in constructing a robust training set for our abstractive Urdu text summarization model. This section provides a detailed account of the steps taken to collect, validate, and preprocess the datasets used for training and fine-tuning. Our primary dataset consists of 67,000 Urdu news articles sourced from Hugging Face. This dataset includes summaries generated using GPT-3.5, which were subsequently validated by human reviewers. Given the challenges faced in this domain—primarily the lack of sufficient summarized datasets for Urdu text—we contacted the dataset creator to understand the construction process. The creator reassured us of the dataset's reliability and provided detailed insights into its creation and validation pipeline. To further ensure the reliability of the dataset, we performed an additional round of validation by randomly sampling 200 entries (0.3%) from the dataset. Each sampled entry was read and reviewed for quality, coherence, and relevance. The results reaffirmed the dataset's reliability, giving us confidence in using it for training. However, some preprocessing was required to ensure compatibility with the models we planned to train. Articles exceeding 6,000 characters were removed, as they could not be fully processed by the model due to token limitations and risked causing hallucinations during summarization. This filtering eliminated 3,000 articles, approximately 4.5% of the original dataset, resulting in a final size of 64,000 usable articles and summaries. No additional cleaning was performed, as the summaries had already undergone validation by both the dataset creator and our team. However since our primary dataset could not be completely validated we created an additional dataset of which we validated every single entry and ensured even better summaries would be used for the fine tuning process for this purpose we sourced 1,000 English news articles from Kaggle. These articles were translated into Urdu using the Google Translate API. Poorly translated articles were identified and discarded, reducing the dataset to 875 usable entries (12.5% discarded). GPT-4 was employed to generate summaries for the translated articles. Every single article and its summary were manually reviewed for accuracy and coherence. This meticulous review process ensured the highest level of quality in this "golden dataset." The secondary dataset thus represents a fully validated, small-scale dataset that complements the larger Hugging Face dataset. It provides a benchmark for fine-tuning and testing, offering complete confidence in its reliability and alignment with our abstractive summarization objectives.

Dataset	Entries (Original)	Final Entries	Discarded (%)	Summary Source	Validation
Hugging Face Dataset	67,000	64,000	4.5%	GPT-3.5 + Human Validation	Randomly sampled 0.3% (~200 entries)
Kaggle Translated Dataset	1,000	875	12.5%	GPT-4 (Summaries) + Manual Validation	Fully validated (100%)

Fig. 2. Dataset Summary and Key Statistics

B. Justification for Dataset Selection

The Hugging Face dataset provides unparalleled coverage and diversity, with summaries validated for reliability through multiple layers of review. It addresses a significant gap in the availability of large-scale Urdu summarization datasets. The secondary dataset, derived from Kaggle, was meticulously validated to ensure complete faith in its quality. The combination of the two datasets enables robust training and fine-tuning, balancing scalability and precision. By relying on the larger Hugging Face dataset for initial training and the smaller, fully validated Kaggle dataset for fine-tuning, we ensure the model learns from diverse, reliable sources while maintaining a high standard of quality for evaluation.

V. Models and Experiments

To effectively tackle the task of abstractive text summarization for Urdu, we required models designed for encoder-decoder tasks, particularly ones optimized for abstractive summarization. Our research began with an extensive evaluation of potential models, prioritizing their applicability to the Urdu language and their suitability for abstractive summarization. We initially considered a range of models, including mBART, mT5, Pegasus, Llama, IndicBART, BERT, and RoBERTa. However, most of these models were deemed unsuitable for our task due to specific limitations:

- **Llama, BERT, and RoBERTa:** These are encoder-only models, making them incompatible with abstractive summarization tasks, which require a decoder component.
- **Pegasus:** While purpose-built for abstractive summarization, Pegasus lacks pretraining on Urdu data, rendering it ineffective for this low-resource language.
- **IndicBART:** Despite its multilingual focus, IndicBART failed to produce high-quality abstractive summaries in Urdu due to insufficient training on the language.
- **mBART and mT5:** These models stood out as they are specifically designed for abstractive summarization and include pretraining on Urdu datasets, making them ideal for our task.

We conducted initial experiments with Llama and IndicBART, but their results confirmed their unsuitability: Llama consistently generated extractive rather than abstractive summaries, which was not the purpose of our research, while IndicBART produced summaries in Hindi with poor coherence as it was not particularly trained on Urdu. Based on these observations, we focused our efforts on fine-tuning mBART and mT5, as these models are best suited for abstractive summarization in Urdu.

Model Comparison	Abstractive Capability	Trained on Urdu	Architecture	Performance
mBart	Yes	Yes	Encoder-Decoder	Selected
mT5	Yes	Yes	Encoder-Decoder	Selected
Pegasus	Yes	No	Encoder-Decoder	Not Suitable
Llama	No	No	Encoder Only	Not Suitable
IndicBART	Partial	Partial	Encoder-Decoder	Not Suitable
BERT, RoBERTa	No	No	Encoder Only	Not Suitable

Fig. 3. Existing Models for Abstractive Summarization

A. Training Parameters and Setup

The models were fine-tuned on a dataset of 64,000 high-quality Urdu news articles and their summaries. Due to tokenization limits, the dataset size was adjusted for each model:

- **mBART:** Processes sequences up to 1024 tokens, allowing it to use the full dataset.
- **mT5:** Limited to 512 tokens, requiring further preprocessing. To optimize performance and minimize hallucination, the training dataset was reduced to 50,000 units for this model.

The following hyperparameters were used during training:

Hyper parameters	mBart	mT5
Epochs	3	2
Batch Size	4	4
Learning Rate	5e-5	3e-5
Gradient Accumulation Steps	4	4
Weight Decay	0.01	0.01
Mixed Precision	fp16	bf16
Evaluation Strategy	Epoch	Epoch

Fig. 4. Models Parameters

The models were trained on NVIDIA A100 GPUs, with each epoch taking approximately 2 hours for mBART and 1 hour and 48 minutes for mT5. These settings were chosen to balance computational efficiency with the model's ability to learn effectively.

B. Evaluation Criteria

We used BERTScore to evaluate the quality of generated summaries. This metric was selected because it measures semantic similarity by comparing embeddings of the predicted and reference summaries, unlike traditional metrics like ROUGE, which rely on surface-level overlaps and fail to capture meaning.

	Best	Worst	Average
mBart	0.622	0.399	0.497
mt5	0.471	0.272	0.355

Fig. 5. BERTScores of mBART and mT5

- **Interpretation:** While the scores indicate room for improvement, the summaries generated by both models are semantically meaningful and qualitatively better than previous extractive approaches. mBART outperformed mT5, producing more coherent and fluent summaries.
- **Comparison with Prior Work:** Previous research primarily used ROUGE, which fails to evaluate the abstractive quality of summaries. Additionally, most prior work generated extractive summaries, making our results a significant improvement.

C. Validation and Manual Review

To ensure the reliability of the generated summaries, we conducted:

1) **Peer Validation:** Randomly selected summaries were reviewed by experts fluent in Urdu, confirming their semantic accuracy and coherence.
2) **Qualitative Assessment:** Examples of generated summaries demonstrate that the outputs are truly abstractive, capturing the essence of the original text while rephrasing and restructuring content.

VI. Results and Discussions

Following the fine-tuning experiments discussed earlier, we observed significant improvements in the performance of the mBART and mT5 models for Urdu abstractive summarization. Our initial experiments with mBART revealed a key issue: the generated summaries often contained excessive repetition of a single sentence.

مختار مائی کا نام رابعہ ہے اور شازیہ کا نام مختار مائی ہے۔ مختار مائی کا نام
رابعہ ہے اور شازیہ کا نام مختار مائی ہے۔ مختار مائی کا نام رابعہ ہے اور شازیہ
کا نام مختار مائی ہے۔ مختار مائی کا نام رابعہ ہے اور شازیہ کا نام مختار مائی
ہے۔ مختار مائی کا نام مختار مائی ہے۔ مختار مائی کا نام مختار مائی ہے۔

Fig. 6. mBart initially trained on 3000 articles

This repetition suggested that the model was not generating optimal summaries, prompting us to make adjustments to certain parameters, namely the following:

- **Number of beams (`num_beams`):** In beam search, the number of beams controls how many different sequences the model explores during decoding. By increasing the number of beams, we can improve the diversity and fluency of the generated summaries.

- **N-gram size:** Adjusting the n-gram size helps reduce repetitive patterns in the summaries by discouraging the model from using the same n-grams excessively.
- **Batch size:** We reduced the batch size from 16 to 4 to avoid memory limitations and to allow for more stable and effective training.
- **Length penalty:** We introduced a length penalty to encourage the model to generate summaries of optimal length, avoiding excessively short or overly long outputs.

After fine-tuning these parameters, we retrained the mBART model on the complete 64k Urdu news dataset and the golden dataset, and the resulting summaries showed marked improvement in terms of coherence, fluency, and contextual relevance.

ایف بی آئی کے ڈائریکٹر کامی نے منگل کے روز سینیٹ کی سماعت میں بتایا کہ روسی ہیکرز نے ریپبلکن نیشنل
کمیٹی کے کمپیوٹر ریکارڈ میں داخل ہو کر اسے "پرانے آر این سی سی" کمپیوٹر سسٹمز کا نام دیا، جو نہیں
تھے۔ کامی کا یہ بیان اس بات کے لئے اہم تھا کہ کمیٹی نے ہفتوں پہلے سائبرسیکیوریٹی کی مضبوطی کی وجہ
سے روسیوں کے ساتھ ڈیٹا نہیں کھوا۔ ٹرمپ نے اس کو دہرایا اور دعویٰ کیا کہ ڈیموکریٹک نیشنل کمیٹی
سسٹم میں کمزوریوں نے ان کے سسٹم کو ہیک کرنے کا راستہ کھول دیا ہے۔ کامی کے مطابق، یہ ڈیٹا روسی
ٹھیکیدار کی جانب سے ایک حملے کا حوالہ دیتے ہوئے آیا، لیکن اس کا کوئی ثبوت نہیں ملا۔ انٹیلیجنس
کمیونٹی نے اطلاع دی کہ روسی تنظیموں نے حملہ کیا تھا، لیکن انہوں نے اس ڈیٹا کو عوامی بنانے کا انتخاب
نہیں کیا تھا۔ رپورٹ میں اس کے بارے میں کوئی خاص حوالہ نہیں دیا گیا۔ اس معاملے پرجمہوریہ مائن کے
سینیٹر سوسن کولنز نے اس بات پر زور دیا کہ ریاستی سطح پر ہیکنگ کی ہدایت یا ہیکنگ کے ثبوت
موجود ہیں۔

Fig. 7. Final-Tuned mBart Result

On the other hand, our work with mT5 initially faced technical challenges. The model failed to run properly with the `fp16` precision setting, which is commonly used for faster training but can cause numerical stability issues on some models. Switching to `bf16` precision, which uses a broader range of numbers with lower precision, allowed the training to proceed smoothly and resolved the issue. This change enabled us to continue fine-tuning the mT5 model effectively.

یہ آپ کو سمجھانے کی بات ہے۔

Fig. 8. mT5 initially training Result

Additionally, we encountered another challenge with mT5: the initial summaries generated were too short. To address this, we introduced the `min_length` parameter with a value of 30, ensuring that the model produced sufficiently detailed summaries. We also adjusted the learning rate to 3e-5 to optimize the training process. Given that both the training loss and validation loss were relatively low, we chose to train the model for 2 epochs to avoid overfitting and ensure efficient convergence. Despite these adjustments, the performance of mT5 was still somewhat inconsistent. While some summaries were relatively coherent and well-structured, others lacked fluency and exhibited issues with sentence clarity.

After training, we compared the summaries generated by both models. The mBART-generated summaries consistently outperformed those from mT5 in terms of fluency and coherence. However, mT5 did produce some sum-

امریکی ریاست نیویارک میں ایک سابق این
ایف ایل ایل پلیئر پر نسلی تعصب کے الزام
میں فرد جرم عائد کر دی گئی ہے۔

Fig. 9. Final-Tuned mT5 Result

maries that were quite good, showing potential but also revealing variability in its performance.

To assess the quality of the generated summaries, we relied on BERTScore, which measures semantic similarity between the predicted and reference summaries. The average BERTScore for mBART was 0.497, while mT5 scored 0.355. These scores highlight the relative strengths of mBART over mT5, but BERTScore alone does not fully capture the quality of the summaries. Summarization is inherently subjective, and factors like fluency, coherence, and contextual relevance play a significant role in evaluating the quality of an abstractive summary. To complement the automated metrics, we also conducted a manual review of the generated outputs. The review revealed that both models produced semantically accurate summaries, and the results were generally very good, with mBART's outputs being more fluent and coherent overall.

In comparison to prior work on Urdu abstractive summarization, which often relied on extractive methods or models that were not large-scale multilingual models, our results with mBART represent a substantial improvement. Previous research in this domain generally did not utilize models like mBART, and the high-quality summaries we were able to generate demonstrate the effectiveness of this model for low-resource languages like Urdu.

VII. Future Work and Recommendations

To improve the model performance, the main emphasis will be made on increasing the training dataset. Since it is difficult to work with low-resource languages such as Urdu, it is crucial to enhance the quantity and variety of high-quality datasets. The difficulty in obtaining large, high-quality Urdu datasets remains a significant hurdle, as such resources are often limited and require considerable manual effort to curate. However, gathering a larger and more diverse dataset will allow the model to better handle various domains and improve its robustness.

In addition to data expansion, optimizing hyperparameters will remain a key avenue for performance improvement. Experimenting with hyperparameters such as learning rate, batch size will help refine the model's ability to generate high-quality, fluent summaries.

Given the computational demands of training large language models, access to high-performance GPUs will be crucial. GPUs will enable more efficient fine-tuning, allowing for faster iteration and the ability to handle larger datasets. This computational capacity will be vital for scaling up experiments and improving the model's overall performance. Ultimately, addressing these challenges will allow us to significantly advance the capabilities of Urdu abstractive summarization models.

VIII. References

[1] A. Raza, H. S. Raja, and U. Maratib, ”Abstractive Summary Generation for the Urdu Language,” 2024 International Conference on Computing, Information Systems & Industrial Engineering, pp. 1-8, 2024.

[2] A. Raza, M. H. Soomro, S. Salahuddin, I. Shahzad, and S. Batool, ”Abstractive Text Summarization for Urdu Language,” Journal of Computing & Biomedical Informatics, vol. 7, no. 2, pp. 1-15, 2024. doi: 10.56979/702/2024.

[3] N. Shafiq, I. Hamid, M. Asif, Q. Nawaz, H. Aljuaid, and H. Ali, ”Abstractive Text Summarization of Low-Resourced Languages Using Deep Learning,” PeerJ Computer Science, vol. 9, pp. e1176, 2023. DOI: 10.7717/peerj-cs.1176.

[4] M. Awais and R. M. A. Nawab, ”Abstractive Text Summarization for the Urdu Language: Data and Methods,” IEEE Access, vol. 12, pp. 1-12, 2024. doi: 10.1109/ACCESS.2024.3378300.

[5] H. Raza and W. Shahzad, ”End to End Urdu Abstractive Text Summarization with Dataset and Improvement in Evaluation Metric,” IEEE Access, vol. 12, pp. 1-12, 2024. doi: 10.1109/ACCESS.2024.3377463.