

# **Optimizing Scientific Paper Summarization with Fine-Tuned T5 on the ArXiv Dataset**

Muhammad Ibtisam Afzal  
FA22-BCS-073  
05 Oct 2024

## **Abstract**

In today's fast-paced scientific environment, the rapid growth of research publications has created a demand for efficient paper summarization tools. The motivation for this study stems from the need to help researchers quickly comprehend large volumes of scientific literature. A key challenge lies in developing summarization models that maintain both coherence and accuracy, despite the complexity of the source material. Previous approaches, including various transformer-based models like T5, have demonstrated potential, but they often struggle with domain-specific nuances and scalability.

In this work, we fine-tune the T5-small model on the ArXiv-summarization dataset to optimize scientific paper summarization. Our contributions include improving the model's ability to recognize critical concepts and structure within scientific texts, enhancing the summarization quality while maintaining computational efficiency. We also employ the ROUGE score for a rigorous evaluation of model performance, highlighting areas for further refinement.

The significance of this research lies in its potential to substantially reduce the time researchers spend reviewing literature, fostering more efficient knowledge dissemination. By advancing summarization capabilities, our model can contribute to the development of scalable, automated tools that streamline the research process across various domains.

## **Keywords**

Scientific paper summarization, Fine-tuned T5 model, Automated summarization, ArXiv Summarization dataset, ROUGE score evaluation, Research efficiency, Scientific Paper Summarization, Abstractive Summarization.