



**Department of Electrical and Computer Engineering
North South University**

Senior Design Project

**Are abstracts really the abstracts: alignment in academic research papers and
summarizing medical research paper with deep learning**

Tanvir Khan ID# 1911481642

Faculty Advisor:

Dr. Mohammad Ashrafuzzaman Khan

Assistant Professor

ECE Department

Spring, 2023

LETTER OF TRANSMITTAL

June, 2023

To

Dr. Rajesh Palit

Chairman,

Department of Electrical and Computer Engineering

North South University, Dhaka

Subject: Submission of Capstone Project Report on “Are abstracts really the abstracts: alignment in academic research papers and summarizing medical research paper with deep learning”

Dear Sir,

With due respect, we would like to submit our **Capstone Project Report** on “**Are abstracts really the abstracts: alignment in academic research papers and summarizing medical research papers with deep learning**” as a part of our BSc program. The report deals with abstractive summarization of medical research papers. This project was very much valuable to me as it helped me gain experience from practical field and apply in real life. I tried to the maximum competence to meet all the dimensions required from this report.

We will be highly obliged if you kindly receive this report and provide your valuable judgment. It would be our immense pleasure if you find this report useful and informative to have an apparent perspective on the issue.

Sincerely Yours,

.....

Tanvir Khan

ECE Department

North South University, Bangladesh

APPROVAL

Tanvir Khan (ID #1911481642) from the Electrical and Computer Engineering Department of North South University, has worked on the Senior Design Project titled “Are abstracts really the abstracts: alignment in academic research papers and summarizing medical research papers with deep learning” under the supervision of Dr. Mohammad Ashrafuzzaman Khan partial fulfillment of the requirement for the degree of Bachelors of Science in Engineering and has been accepted as satisfactory.

Supervisor’s Signature

.....
Dr. Mohammad Ashrafuzzaman Khan

Assistant Professor

Department of Electrical and Computer Engineering

North South University

Dhaka, Bangladesh.

Chairman’s Signature

.....
Dr. Rajesh Palit

Professor

Department of Electrical and Computer Engineering

North South University

Dhaka, Bangladesh.

DECLARATION

This is to declare that this project is our original work. No part of this work has been submitted elsewhere partially or fully for the award of any other degree or diploma. All project-related information will remain confidential and shall not be disclosed without the formal consent of the project supervisor. Relevant previous works presented in this report have been properly acknowledged and cited. The plagiarism policy, as stated by the supervisor, has been maintained.

Students' name & Signature

Tanvir Khan

ACKNOWLEDGEMENTS

The author would like to express their heartfelt gratitude towards their project and research supervisor, Dr. Mohammad Ashrafuzzaman Khan, Assistant Professor, Department of Electrical and Computer Engineering, North South University, Bangladesh, for his invaluable support, precise guidance and advice pertaining to the experiments, research and theoretical studies carried out during the course of the current project and also in the preparation of the current report.

Furthermore, the authors would like to thank the Department of Electrical and Computer Engineering, North South University, Bangladesh, for facilitating the research. The authors would also like to thank their loved ones for their countless sacrifices and continual support.

ABSTRACT

Are abstracts really the abstracts: alignment in academic research papers and summarizing medical research papers with deep learning

This research project investigates the degree of alignment between academic research and its summarizing abstracts in the medical domain, employing deep learning models for abstractive summarization. Abstractive summarization is a technique that generates summaries of texts using natural language generation rather than extracting sentences from the original texts. However, most existing studies on abstractive summarization focus on news articles or short stories, and few have explored its application to scientific literature, especially medical research papers. We used two state-of-the-art deep learning models, bert2bert and roberta2roberta, which are based on the encoder-decoder architecture and use pre-trained language models as encoders and decoders. We trained our models on a dataset of full-text papers and their abstracts from CORA-2019 and other sources, which provided access to a large and diverse collection of medical research papers. We evaluated the performance of our models using ROUGE and METEOR scores, which are commonly used metrics for measuring the quality of summaries. We found that roberta2roberta outperformed bert2bert on both metrics, achieving a ROUGE score of 0.50 and a METEOR score of 0.48. We also compared the generated summaries with the original abstracts and performed a qualitative analysis of their differences and similarities. We found that the abstracts in the papers most often suffer from some common errors and limitations, such as repetition, inconsistency, omission, or exaggeration. We identified some potential benefits and challenges of applying abstractive summarization to scientific literature, such as improving accessibility and readability of research papers, reducing information overload, enhancing scientific communication, promoting interdisciplinary research, ensuring ethical and responsible use of AI, protecting intellectual property rights, preserving data quality and integrity, ensuring data privacy and security, reducing environmental footprint, etc. We concluded abstracts don't always reflect the whole paper through abstractive summarization of medical research papers, but it still requires further research and development to overcome its challenges and limitations.