



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

**Directed Research**

**Fine-tuning Transformer Decoder Model for Disease  
Prediction and Clinical Data Generation**

**Hryddhi Mirza**

**1912854042**

**Faculty Advisor:**

**Dr. Mohammad Ashrafuzzaman Khan**

**Assistant Professor**

**ECE Department**

**Fall 2022**

# LETTER OF TRANSMITTAL

November, 2023

To

Dr. Rajesh Palit

Chairman,

Department of Electrical and Computer Engineering

North South University, Dhaka

**Subject: Submission of Directed Research Report on “Fine-tuning Transformer Decoder Model for Disease Prediction and Clinical Data Generation”**

Dear Sir,

With due respect, I would like to submit my **Directed Research Report** on “Fine-tuning Transformer Decoder Model for Disease Prediction and Clinical Data Generation” as a part of my BSc program. This project was very much valuable to me as it helped us 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.

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

Sincerely Yours,

.....

Hryddhi Mirza

ECE Department

North South University, Bangladesh

# APPROVAL

Hryddhi Mirza (1912854042) from Electrical and Computer Engineering Department of North South University, has worked on the Directed Research Project titled “Fine-tuning Transformer Decoder Model for Disease Prediction and Clinical Data Generation” 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 my 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' names & Signatures

---

**Hryddhi Mirza**

# ACKNOWLEDGEMENTS

The author would like to express her heartfelt gratitude towards the 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 author would like to thank the Department of Electrical and Computer Engineering, North South University, Bangladesh for facilitating the research. The author would also like to thank her loved ones for their countless sacrifices and continual support.

# Abstract

## **Fine-tuning Transformer Decoder Model for Disease Prediction and Clinical Data Generation**

In today's rapidly evolving technological landscape, new scientific breakthroughs are taking place frequently in all sectors. But only general information can be found through a simple internet search using symptoms as keywords for a disease. To avoid such situations, we aim to generate similar or related disease from a given case and associated symptoms in order to help the doctors and medical professionals to aid in patient care using related documents written by researchers from all across the world.

### **1 INTRODUCTION**

The idea is to generate a similar disease based on a given case and few symptoms that are given as input to support the doctors, medical professionals and other healthcare workers in order to provide quality patient care by using the vast information and research documents written by experts from all around the world. To attain this objective, significant amounts of data has been gathered and used from the "*COVID-19 Open Research Dataset in response to the COVID-19 pandemic (CORD-19)*". This dataset represents a vast pool of knowledge that has been compiled and studied by experts who are in the field and underwent rigorous critiques. Our work manages to generate a suggestion for a related disease more accurately than Google, as Google is just a search engine it only gives a recommendation of a few diseases and does not specify a particular disease. Not only that, our work is also better than ChatGPT as ChatGPT generates an answer very confidently even if it is wrong. The broad documentation provided includes in-depth understanding of various medical conditions and their underlying causes, thus promoting a better understanding of diseases.

With the help of this model doctors, medical professionals along with the patients will be able to identify a related and specific disease just by giving a disease as a case and few symptoms as the input, not only that the model will give an idea of what that given disease is, it also guarantees to give the most suitable solution for the given disease, reducing the chances of errors, as the results are generated from validated and precise experiments, which saves significant amount of time which increases the efficiency.