

Department of Electrical and Computer Engineering  
North South University

---



## SENIOR DESIGN PROJECT

CSE499A Capstone  
in  
Building COVID Knowledge Base from Medical  
Journals

Spring 2021

*Faculty Advisor:*

Dr. Mohammad Ashrafuzzaman Khan  
Assistant Professor  
Department of ECE

*Submitted by*

Md. Jubaer Khan ID:1721616042  
Walidul Alam Ehab ID:1631214042

## Declaration

We are pleased to submit our report on the topic — Building COVID Knowledge Base from Medical Journals, we have created is not a clone of any other team's "Capstone Project Report." So, the authenticity of this report is hidden under its one-of-a-kindness because we have written it on our own. And it was only for this reason that it was designed and submitted. It was never used for anything else and would never be used for anything else.



---

Walidul Alam Ehab  
ID:1631214042  
Department of ECE  
North South University, Bangladesh



---

Md. Jubaer Khan  
ID:1721616042  
Department of ECE  
North South University, Bangladesh

## Approval

The Project report titled Building COVID Knowledge Base from Medical Journals has been submitted by Walidul Alam Ehab ID:1631214042 and Md. Jubaer Khan ID:1721616042 to the Department of Electrical & Computer Engineering of North South University, Dhaka, Bangladesh, for the partial fulfillment of the requirements for the Degree of Bachelor of Science (BSc) in Computer Science and Engineering.

### Supervisor's Signature



---

**Dr. Mohammad Ashrafuzzaman Khan**  
**Assistant Professor**

Department of Electrical and Computer Engineering  
North South University  
Dhaka, Bangladesh.

### Department Chair's Signature

---

**Dr. Mohammad Rezaul Bari**  
**Associate Professor and Chairman**

Department of Electrical and Computer Engineering  
North South University  
Dhaka, Bangladesh.

## **Acknowledgment**

First, we would like to thank Almighty Allah, for blessing us with the capability to successfully conducting this project

We wish to express our deepest appreciation to our supervisor, Dr. Mohammad Ashrafuzzaman Khan, for giving us with this research topic , helpful suggestions, supervision and moral encouragement in order to complete this project. Special thanks to everyone who helped us complete this report and provided us with papers to help us with our endeavor.

We would also like to thank to the North South University's ECE department for allowing us to undertake this research. And also want to thank the COVID-19 data-set maker team and research contributor in "COVID-19 Open Research Dataset Challenge"

Finally, we want to express our gratitude to our families and everyone who helped us complete this project by providing support and direction.

## **Abstract**

Natural Language Processing (NLP), a subfield of linguistics, computer science and artificial intelligence, is concerned with the interactions between computers and human language, particularly with how computers are programmed to process and interpret large quantities of natural language data. In the covid pandemic, many research articles are being published every day, and scientists and doctors are trying to figure out the cure for saving humanity. Unfortunately, it is pretty hard to keep updated with all of the research that is being published. Our approach is to extract essential information from the research articles and find a relation between the disease and its symptoms using the Word2vec model. We have preprocessed a large number of journal article's abstracts and train them into Word2vec model in our work. After getting the words vectorial representation, we have visualized the data for finding significant relationships among the disease and their symptoms-related words.