North South University

Department of Electrical and Computer Engineering

School of Engineering and Physical Sciences (SEPS)

Senior Project Design

Transfer Learning for Speaker Diarization on Bangla Audio Dataset

A.N.M Fahim Faisal – 1711758642

Umnoon Binta Ali – 1713013042

Md. Rayhan Talukder – 1712356642

Md. Afikur Rahman – 1711035042

Faculty Advisor

Ms. Tanjila Farah

Senior Lecturer

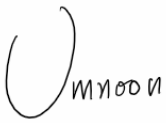
Summer 2021

DECLARATION

We, hereby, declare that the work presented in this report is the outcome of our four months' work performed under the supervision of Ms. Tanjila Farah, Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh. The work was spread over a span of one of the final year courses, CSE 499A and CSE499B, Senior Design Project I and II, in accordance with the course curriculum of the Department for the Bachelor of Science in Electrical and Electronics Engineering program.

Students' name & signature:

A.N.M Fahim Faisal



Umnoon Binta Ali

Md. Rayhan Talukder

Md. Afikur Rahman

Approval

The senior project report on 'Transfer Learning Model in Speaker Diarization on Bangla Audio Dataset' has been submitted by MD. Rayhan Talukder (1712356642), A.N.M Fahim Faisal (1711758642), Md. Afikur Rahman (1711035042), Umnoon Binta Ali (1713013042) students of the Department of Electrical and Computer Engineering, North South University, Bangladesh. This report partially fulfills the requirement for the degree of Bachelor of Science in Electrical and Electronics Engineering in December 2019 and has been accepted as satisfactory.

Supervisor's Signature
Ms. Tanjila Farah
Senior Lecturer
Department of Electrical and Computer Engineering North South University, Dhaka, Bangladesh

Department Chair's Signature
Dr. Mohammad Rezaul Bari
Associate Professor and Chair
Department of Electrical and Computer Engineering North South University, Dhaka, Bangladesh

Acknowledgments

First, we express our gratitude to almighty ALLAH for His blessing which makes us possible to complete the project. We are grateful and wish our profound indebtedness to Ms. Tanjila Farah, Senior Lecturer, Department of ECE, North South University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "Deep Learning" to carry out this project. Her guidance, constant supervision, enthusiastic encouragement, sagacious advice, and effective surveillance throughout the entire period of the project have made it possible to complete this project.

We would like to thank our entire CSE 499 course mate in North South University, who took part in this discussion while completing the course work. At last, we must express our sincere heartfelt gratitude to all the staff members of the Computer Engineering Department who helped us directly or indirectly during this course of work.

Abstract

Sound classification is a very intriguing concept in the field of artificial intelligence. Speaker Diarization is a very interesting task in the domain of sound classification. It has recently expanded significantly with the introduction of deep learning technology, which has transformed research and practices throughout speech application fields. Speaker diarization is the process of assigning labels to audio data that match the speaker's identity. It is quite beneficial when it comes to identifying audio information. In Bangla, very little work has been done on speaker diarization. This project aimed to build a Bangla dataset for the diarization process in this research. We described our deep learning model for the speaker diarization problem and demonstrated how transfer learning can be utilized to swiftly learn a model with minimal performance loss when compared to a fully trained one. To increase the universal applicability of our model, we focused on transfer learning and tweaked it manually across AMI and Bangla dataset. Additionally, we've been focusing our efforts on improving the Diarization Error Rate (DER) and experimenting with other embedding generation networks. We obtained a DER score of 0.24 using our transfer learning variation trained on Bangla dataset.