LOW RESOURCE SPEECH INTENT CLASSIFICATION USING MFCC FEATURES

Anas Fathima Rifaza

219393M

Master of Science in Computer Science

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

March 2025

LOW RESOURCE SPEECH INTENT CLASSIFICATION USING MFCC FEATURES

Anas Fathima Rifaza

219393M

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree

Master of Science in Computer Science

Department of Computer Science and Engineering

Faculty of Engineering

University of Moratuwa

Sri Lanka

March 2025

DECLARATION

I hereby declare that this thesis is the result of my own independent work. It does not include any content that has been previously submitted for a degree or diploma at any university or institute of higher education, unless properly cited. To the best of my knowledge, all materials taken from the work of others have been appropriately acknowledged and referenced within the text. I also reserve the right to reuse parts of this work in future publications, such as journal articles or academic books.

Signature: Date:

The above candidate has carried out research for the PhD/MPhil/Masters thesis/dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr.T.Uthayasanker

Signature of the Supervisor: Date: 28 Jul 2025

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my MSc Research Project supervisor, Dr. Uthayasanker Thayasivam, for his invaluable guidance, continuous support, and encouragement throughout my research journey. His expertise and insights have been instrumental in shaping this work, and his unwavering support in providing the necessary resources has greatly contributed to the successful completion of my MSc thesis.

I am also deeply grateful for his constructive feedback, mentorship, and valuable suggestions, which have significantly enhanced the quality of this research. Additionally, I extend my heartfelt appreciation to my colleagues for their assistance in exploring relevant research materials and fostering a collaborative learning environment.

Furthermore, I am immensely thankful to my family—my parents, siblings, nephew, niece, and close friends—for their unwavering encouragement and support throughout this journey. Their belief in me has been a constant source of motivation.

Finally, I would like to extend my appreciation to everyone who has contributed to this endeavor, whether directly or indirectly. Their support has been invaluable in helping me navigate the challenges of my MSc studies

ABSTRACT

Speech-based user interfaces have revolutionized digital interactions, yet developing them for low-resource languages remains a challenge due to limited labeled speech data. This research proposes a Convolutional Neural Network (CNN)-based approach utilizing Mel-Frequency Cepstral Coefficients (MFCC) along with delta and delta-delta features for effective speech intent classification in Sinhala and Tamil. The methodology incorporates audio preprocessing, MFCC feature extraction, and data augmentation techniques such as noise addition, pitch shifting, and time stretching. A stratified cross-validation framework is used to ensure fair and consistent evaluation. The proposed model achieves 96.92% accuracy on the Sinhala dataset (7,624 samples) and 93.81% on the Tamil dataset (400 samples, ~0.5 hours of speech), representing a substantial improvement over prior methods. These results demonstrate the effectiveness of the CNN-based approach in capturing meaningful acoustic patterns for intent recognition in low-resource settings. The study offers a scalable, efficient solution for speech intent classification and contributes to the advancement of inclusive voice-enabled technologies.

Keywords: Speech Intent Classification, Low-Resource Languages, Pre trained Models, Convolutional Neural Network (CNN), Transfer Learning, Mel-Frequency Cepstral Coefficients (MFCC).

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Description** **Page**

LIST OF TABLES

Table **Description** **Page**

LIST OF ABBREVIATIONS

Abbreviation	Description
MFCC	Mel-Frequency Cepstral Coefficients
ASR	Automatic Speech Recognition
NLP	Natural Language Processing
HMMs	Hidden Markov Models
RNNs	Recurrent Neural Networks
NLU	Natural Language Understanding
OOV	Out-of-Vocabulary
SVM	Support Vector Machine
ROC	Receiver Operating Characteristic