Lip Reading Using Deep Learning Techniques

A PROJECT REPORT

Submitted by,

Mr. Mohammed Dhayan Ahmed - 20211CSD0097
Mr. Srivatsa H - 20211LSD0004
Mr. Bavith Raj - 20211CSD0056

Under the guidance of,

Mr. LAKSHMISHA S K

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

At

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY

BENGALURU

MAY 2025

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

This is to certify that the Project report **"Lip Reading using Deep Learning Techniques"** being submitted by "Mohammed Dayan Ahmed, Srivatsa H, Bavith Raj" bearing roll number(s) "20211CSD0097, 20211LSD0004, 20211CSD0056" in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

Mr. Lakshmisha S K
Assistant Professor
School of CSE
Presidency University

Dr. Saira Banu Atham
Prof. & HoD
School of CSE
Presidency University

14/5/25

Dr. Mydhili Nair
Associate Dean
School of CSE
Presidency University

Dr. Sameeruddin Khan
Pro-VC School of Engineering
Dean -School of CSE
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **Lip Reading using Deep Learning Techniques** in partial fulfillment for the award of Degree of Bachelor of **Technology** in **Computer Science and Engineering**, is a record of our own investigations carried under the guidance of **Mr. Lakshmisha SK, School of Computer Science Engineering, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

NAME	ROLL NO	SIGNATURE
MOHAMMED DHAYAN AHMED	20211CSD0097	
SRIVATSA H	20211LSD0004	
BAVITH RAJ	20211CSD0056	

ABSTRACT

The increasing need for natural human-computer communication and access tools has driven the advancement of automated lip-reading systems that can interpret speech based on visual inputs alone. Handcrafted feature-based methods and rule-based classifiers have failed to deliver consistent performance across varied real-world settings because of speaker appearance, lighting, and articulation variability. As a response, this work suggests a deep learning-driven lip-reading system using Convolutional Neural Networks (CNNs) for spatial feature learning and Gated Recurrent Units (GRUs) for temporal sequence modeling.

A special dataset was designed with short video segments having single spoken words. The clips were preprocessed to obtain mouth regions, frame conversion to grayscale, and input dimension normalization for training the model. The CNN-GRU model was trained to predict sequences of lip movements into word classes to make real-time speech prediction from webcam input as well as from uploaded videos.

The system is also strengthened by incorporating a preprocessing pipeline with MediaPipe for stable mouth detection, and by utilizing collate functions to accommodate variability in the number of frames. The proposed architecture achieves high word-level accuracy in both real-time and offline inference modes. Experimental results show that the model generalizes well across users and performs stably under moderate variations in lighting.

Compared to traditional visual speech recognition systems, this deep learning method has a number of benefits, such as enhanced accuracy, scalability, and flexibility in adapting to unknown users. The end-to-end trainable architecture also minimizes reliance on manual feature engineering. The fact that the system can predict spoken words without audio input makes it especially beneficial for applications in accessibility, silent communication, and security.

This project adds to the area of visual speech recognition with a real-world, scalable, and accurate solution for lip reading through contemporary deep learning methods. Vocabulary expansion, multimodal fusion with audio input, and application in real-world assistive technology will be examined in future work.

ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC, School of Engineering and Dean, School of Computer Science and Engineering, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr. Mydhili Nair**, School of Computer Science and Engineering, Presidency University, and **Dr. Saira Banu Atham** Head of the Department, School of Computer Science and Engineering, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Mr. Lakshmisha S K** Assistant Prof, School of Computer Science and Engineering, Presidency University for her inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the CSE7301 Capstone Project Coordinators **Dr. Sampath A K and Mr. Md Zia Ur Rahman,** Git hub coordinator **Mr. Muthuraj.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Mohammed Dhayan Ahmed - 20211CSD0097

Srivatsa H – 20211LSD0004

Bavith Raj – 20211CSD0056