**DEEP LEARNING PACKAGE – ABSTRACT**

Ears of Tomorrow : A Symphony for Hearing Impaired

**TEAM MEMBERS**

Dhivyadharshini B (21PD06)

Harshini T S (21PD12)

E Jagadeesh (21PD13)

**AIM**

To develop Automatic Speech Recognition (ASR) tool for both Audio and Video that could be helpful for Hearing Impaired Individuals.

**ABOUT THE PACKAGE**

**BASED ON VIDEO :**

As human lipreading performance is poor. Hearing-impaired people achieve an accuracy of only 17±12% even for a limited subset of 30 monosyllabic words and 21±11% for 30 compound words. Lipreading is the task of decoding text from the movement of a speaker's mouth. Traditional approaches separated the problem into two stages: designing or learning visual features, and prediction. The developed model LipNet, maps a variable-length sequence of video frames to text. This model uses the GRID corpus dataset for training as well as testing the model.

LipNet is a neural network architecture for lipreading that maps variable-length sequences of video frames to text sequences, and is trained end-to-end. Static slicing was done and the loss used was CTC loss. Then the model was tested and the Speech was finally converted to text.

**BASED ON AUDIO :**

This was done based on two models: Whisper and deepspeech. Whisper is an ASR system trained on 680,000 hours of multilingual and multitask supervised data from the web. The Whisper architecture is implemented as an encoder-decoder Transformer. This model is used for both Trascription and Translation. Deepspeech is a speech-to-text (STT) model developed by Mozilla that converts spoken language into written text. This architecture involves both CNN and RNN. The model is trained on a large corpus of speech data to accurately recognize and interpret various spoken languages and accents.