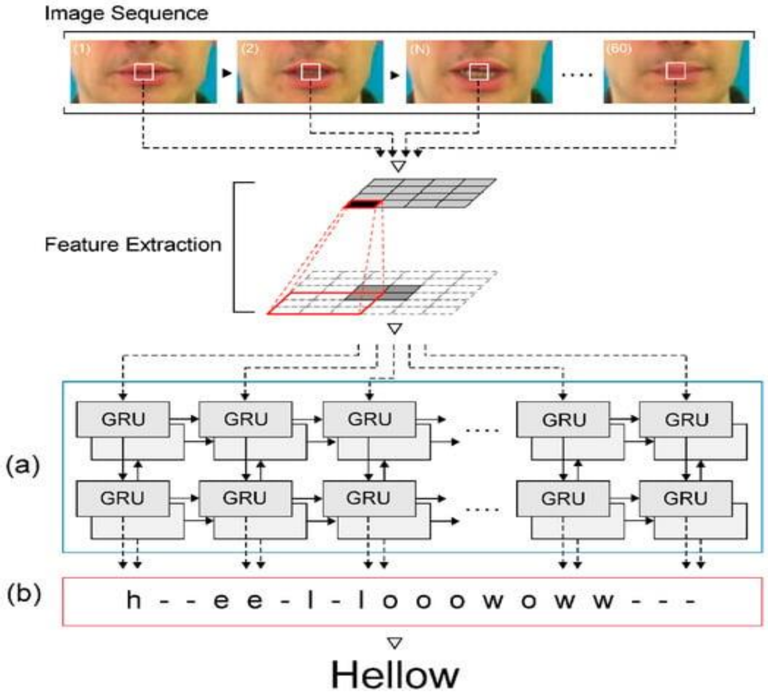
Pro006Aect Design Phase-II
Technology Stack (Architecture & Stack)

Date	20 October 2023
Team ID	591865
Project Name	Project – lip reading using deep learing
Maximum Marks	4 Marks

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2



Reference: https://www.mdpi.com/sensors/sensors-22-00072/article_deploy/html/images/sensors-22-00072-g007-550.jpg

Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	Web UI for user interaction built with streamlit	Streamlit , Vscode.
2.	Video Preprocessing Logic	The video preprocessing logic is responsible for preparing the raw video data for input into the lip-reading model. This includes loading video files, extracting frames, converting them to grayscale, and cropping to focus on the lip region. The goal is to transform raw video data into a format suitable for further analysis by the lip-reading model.	• OpenCV: For loading and processing video data. • TensorFlow: For image transformations and normalization. • NumPy: For numerical operations and array manipulation.
3.	Alignment and Tokenization Logic	This logic handles the alignment of video frames with corresponding transcriptions and tokenizes the transcriptions into numerical sequences for model input. Proper alignment ensures that each frame corresponds to the correct part of the transcription, and tokenization converts textual information into a format the model can understand.	• A pre-trained alignment model or algorithm: For aligning video frames with transcriptions. • TensorFlow or PyTorch: For tokenizing transcriptions using the character-to-number mapping. • Custom scripts: To handle padding or truncation of sequences for uniform input sizes.
4.	Model Prediction and Post-Processing Logic	This logic involves feeding the preprocessed data into the lip-reading model, obtaining predictions, and post-processing the results to generate meaningful output. It includes decoding algorithms to convert model outputs into readable text and additional post-processing steps to	Lip-reading model (TensorFlow, PyTorch): The trained deep learning model for lip reading. CTC Decoding: Connectionist Temporal Classification decoding algorithms for converting model

		handle formatting and language-specific nuances.	outputs into readable text. Python scripting: For post-processing steps, handling spaces, special characters, and language-specific nuances.
5.	Database	Manages data types, configurations, and storage for lip-reading application.	MySQL, NoSQL.
6.	File Storage	Manages file storage requirements for the lip-reading application. This includes storing preprocessed video frames, alignment information, and other necessary data	Local Filesystem
7.	Deep Learning Model	he model contributes to making information accessible to a broader audience, including individuals with hearing impairments, by providing an alternative means of understanding spoken language.	TensorFlow OpenCV NumPy Matplotlib (Pyplot) gdown

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	• TensorFlow • OpenCV • NumPy • Matplotlib (Pyplot) • gdown	• Deep Learning Library • Computer Vision • Numerical Computing • Data Visualization • File Downloading
2.	Scalable Architecture	• A three-tier architecture enables scalable lip-reading by separating user interface, application logic, and data storage for independent scaling.	• Docker, Kubernetes
3	Performance	Optimized Video Processing, Batch Processing, Asynchronous Processing	• Redis or in-memory caching • TensorFlow batch processing • OpenCV

References:

- Easton, R.D.; Basala, M. Perceptual dominance during lipreading. *Atten. Percept. Psychophys.* 1982, *32*, 562–570. [Google Scholar] [CrossRef] [PubMed][Green Version]

- Chung, J.S.; Senior, A.; Vinyals, O.; Zisserman, A. Lip reading sentences in the wild. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 3444–3453. [Google Scholar]

- Kastaniotis, D.; Tsourounis, D.; Fotopoulos, S. Lip Reading Modeling with Temporal Convolutional Networks for Medical Support applications. In *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*;

IEEE: Chengdu, China, 2020; pp. 366–371. [Google Scholar]