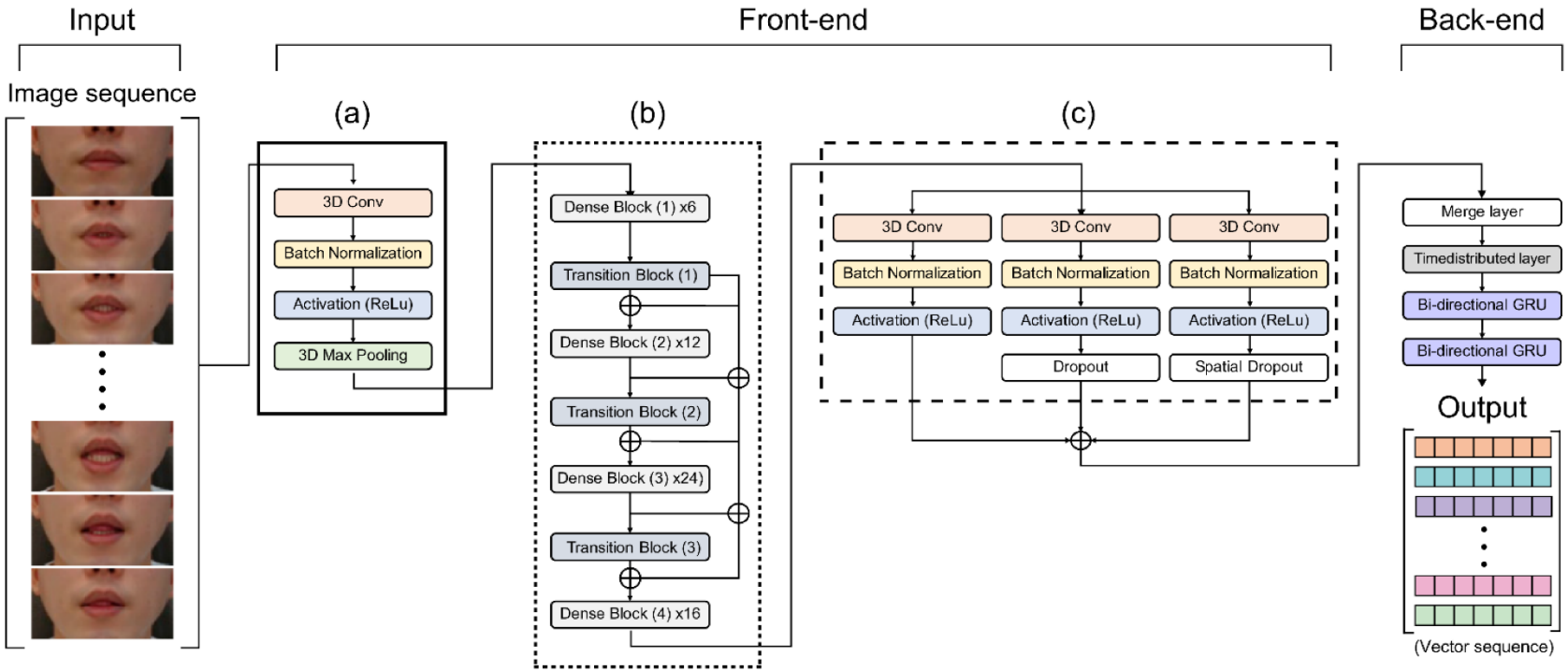
Project Design Phase-II
Technology Stack (Architecture & Stack)

Date	18 November 2023
Team ID	Team-591977
Project Name	Lip Reading using Deep Learning
Maximum Marks	4 Marks

Technical Architecture:



S.No	Component	Description	Technology
1.	User Interface	1.Video Input: - The UI allows users to input video data, either from a live camera feed or pre-recorded video files. - It may include functionalities like starting/stopping video capture, selecting input sources, and adjusting settings. 2. Processing & Transcription: - Upon receiving video input, the system processes frames through the lip reading model. - The UI displays the transcription or translated text generated by the model in real-time or after processing. 3. User Feedback and Controls: - Users might have options to provide feedback, correct transcriptions, or interact with the system to improve accuracy. - Controls for initiating processing, stopping, or pausing the video analysis may be included. 4. Visual Output: - The UI may present the video feed alongside the transcribed text or translation for user reference	1. Front-end Development: - Frameworks/Libraries:HTML, CSS, JavaScript, React, Angular, or Vue.js for creating interactive and responsive user interfaces. - Video Capture: Web APIs (such as getUserMedia) for accessing camera feeds in web applications. 2. Backend & Machine Learning Integration: - Programming Languages:Python for building the machine learning model and backend processing. - Machine Learning Frameworks: TensorFlow, PyTorch, Keras for training and deploying lip reading models. - APIs and Communication:RESTful APIs or GraphQL for communication between the front-end and backend components.
2.	Application Logic-1	Lip reading with machine learning involves the use of algorithms and models to interpret and understand spoken language by analyzing the	1. Deep Learning Models: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are commonly

		movement and shapes of the lips. This technology aims to assist individuals with hearing impairments or be used in scenarios where audio information is not available or reliable.	used for lip reading tasks. These models can capture spatial and temporal features crucial for interpreting lip movements. 2. Dataset:Large datasets containing video clips of people speaking are used to train the models. These datasets are annotated with corresponding transcripts to enable supervised learning. 3. Preprocessing Techniques:Lip images are preprocessed to enhance features, reduce noise, and normalize lighting conditions. Facial landmark detection may be employed to precisely locate key points on the lips. 4. Speech-to-Text Integration: The output of lip reading models is often integrated with speech-to-text technology to convert the interpreted lip movements into written text.
3.	Application Logic-2	Similar to Application Logic-1, this scenario involves the use of machine learning for lip reading. However, the specific application or context may differ. For instance, it could be focused on real-time	1. Real-time Processing: Algorithms optimized for real-time performance are crucial in scenarios where lip reading needs to be performed on the fly.

		lip reading in noisy environments or in conjunction with other sensory inputs.	2. Multimodal Integration: Combining lip reading with other sensory inputs, such as audio or facial expressions, to enhance the accuracy and context of the interpretation.
4.	Application Logic-3	In this scenario, lip reading with machine learning could be applied in a unique context or for a specific industry, such as human-computer interaction, security, or healthcare.	1. Security Applications: Lip reading could be integrated into security systems for identity verification, where lip movements contribute to the authentication process. 2. Healthcare Applications: Lip reading might be used in healthcare for monitoring patients' verbal communication, providing insights into their well-being.
5.	Database	1. Purpose: - Store training data: Databases are used to store video clips or frames of individuals speaking for training the lip reading machine learning models. - Manage metadata: Information about the lip reading models, training parameters, and performance metrics can be stored in the database. - User data: If the application involves user-specific information or preferences, the database may store such data.	1. MySQL: - Use Case: MySQL can be suitable for structured data, and it is commonly used for managing metadata and structured information related to lip reading applications. - Advantages: ACID-compliant, strong community support, and well-suited for relational data. 2. NoSQL (e.g., MongoDB):

		2. Data Type: - Video frames: Individual frames of lip movements, possibly annotated with corresponding transcripts. - Model parameters: Parameters and configurations of the machine learning models. - User profiles: If the application has user-specific features, the database may store user profiles. 3. Configurations: - Scalability: The database should be scalable to handle a potentially large volume of video data, especially during the training phase. - Reliability: Given the critical nature of data for training models, the database should ensure data integrity and reliability.	- Use Case: NoSQL databases are suitable for handling unstructured data, making them a good choice for storing video frames and other non-relational data in lip reading applications. - Advantages: Scalability, flexibility in handling diverse data types, and faster development cycles. 3. IBM Cloudant: - Use Case: Cloudant is a distributed database service that can be beneficial for applications deployed in the cloud, providing scalability and availability. - Advantages: Fully managed, JSON-based document store, and designed for high availability. 4. Graph Database (e.g., Neo4j): - Use Case: If relationships between different elements in the lip reading data are crucial, a graph database may be suitable. - Advantages: Efficient for querying complex relationships, suitable for certain types of analytical tasks.
6.	Cloud Database	For lip reading by machine learning, a cloud database service is used to store and manage data related to lip movements, video clips, and other relevant information. Cloud databases	The technology chosen for the cloud database could be IBM Db2 on IBM Cloud, IBM Cloudant, Amazon DynamoDB, or another suitable cloud-

		provide scalability, accessibility, and reliability, crucial for handling large datasets used in machine learning models for lip reading	based database service. The choice may depend on factors such as performance requirements, scalability, and integration capabilities with other components of the system
7.	File Storage	File storage is essential in lip reading applications for storing video clips, images, and other data related to lip movements. This component is responsible for managing the storage requirements efficiently to support the machine learning models and other processes involved in lip reading.	The file storage technology can include IBM Block Storage, which provides scalable and high-performance block storage on the cloud. Alternatively, other storage services compatible with the chosen cloud platform can be used. Additionally, the application might utilize the local filesystem for temporary storage or caching, depending on specific requirements. The choice of technology would consider factors such as data access speed, scalability, and cost-effectiveness.
8.	External API-1	External API-1, which is the IBM Weather API, is utilized in the lip reading application to provide real-time weather information. While the primary focus of the application is on lip reading through machine learning, weather information may enhance the context or functionality. For example, the application might use weather data to adapt its behavior based on environmental conditions	The IBM Weather API is the chosen technology for accessing weather data. This API allows the application to retrieve up-to-date weather information, which can be used in conjunction with lip reading outputs to provide a more context-aware user experience.
9.	External API-2	External API-2, which is the Aadhar API, is integrated into the lip reading application for	The Aadhar API is the technology chosen for interfacing with the Aadhar

		specific purposes related to identity verification or user authentication. The Aadhar API might be used to cross-verify user identity based on Aadhar information, adding an additional layer of security or personalization to the application.	authentication system. This API allows the application to securely access and verify user identity information, contributing to the overall functionality and security of the lip reading application.
10.	Machine Learning Model	The machine learning model employed in the lip reading application serves the purpose of object recognition, specifically recognizing and interpreting lip movements from video or image data. The model is trained to understand the visual cues and patterns associated with spoken language, contributing to the overall lip reading functionality of the application.	The technology associated with the machine learning model would include frameworks and libraries for object recognition. This could involve deep learning frameworks like TensorFlow or PyTorch, and specific pre-trained models or custom models designed for lip reading tasks.
11.	Infrastructure (Server / Cloud)	The infrastructure component is responsible for deploying the lip reading application, and it offers flexibility in terms of deployment options. The application can be deployed either on a local system or on cloud-based services, providing scalability and accessibility to a wider audience.	The technology associated with infrastructure deployment includes options such as deploying locally on specific servers, using cloud platforms like Cloud Foundry, or utilizing container orchestration systems like Kubernetes. The choice of technology depends on factors such as scalability requirements, resource availability, and deployment preferences. Local deployment may be suitable for smaller-scale usage, while cloud or containerized solutions offer scalability and accessibility benefits.

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Open-source frameworks play a crucial role in the development of the lip reading application. These frameworks provide tools and libraries for machine learning, computer vision, and other essential tasks. TensorFlow, PyTorch, and OpenCV are examples of open-source frameworks commonly used in machine learning applications for tasks like lip reading.	The technology associated with open-source frameworks includes the specific features and functionalities provided by these frameworks. For example, TensorFlow and PyTorch offer deep learning capabilities, while OpenCV provides computer vision tools. The technologies within these frameworks contribute to the development and training of machine learning models for lip reading.
2.	Security Implementations	Security implementations in the lip reading application ensure the confidentiality, integrity, and availability of sensitive data. This includes encryption mechanisms, access controls, and adherence to security best practices such as those defined by OWASP (Open Web Application Security Project).	The technology associated with security implementations includes cryptographic algorithms like SHA-256 for hashing, encryption techniques to secure data in transit and at rest, Identity and Access Management (IAM) controls to manage user access, and adherence to OWASP guidelines for web application security.
3.	Scalable Architecture	The architecture of the lip reading application is designed to be scalable, ensuring that it can handle increasing workloads efficiently. This scalability can	The technology associated with a scalable architecture includes the use of microservices, which involves

		be achieved through architectural choices such as microservices and containerization	breaking down the application into smaller, independent services. Containerization technologies like Docker and orchestration tools like Kubernetes can be employed to manage and scale these services. Load balancers may also be used to distribute incoming traffic across multiple servers, ensuring optimal resource utilization

S.No	Characteristics	Description	Technology
4.	Availability	Ensuring high availability is crucial for the lip reading application, as it needs to be accessible to users consistently. Load balancers are employed to distribute incoming traffic across multiple servers, preventing any single point of failure. The use of distributed servers contributes to redundancy and fault tolerance, enhancing the overall availability of the application.	The technology associated with ensuring availability includes the implementation of load balancers, which intelligently distribute incoming requests across multiple servers. Distributed server configurations, possibly in a cloud environment, enhance fault tolerance and ensure the application remains available even in the event of server failures.
5.	Performance	Performance considerations are essential for a lip reading application to deliver a responsive and efficient user experience. Design considerations include optimizing for the number of requests per second, utilizing caching mechanisms to store	The technology associated with performance optimization includes monitoring tools to analyze the number of requests per second and identify potential bottlenecks. Caching

		frequently accessed data, and employing Content Delivery Networks (CDN) for efficient content delivery, especially in scenarios with high user demand.	technologies enhance response times by storing and serving frequently requested data quickly. CDNs distribute static content geographically to reduce latency and improve overall performance. The choice of these technologies depends on the specific requirements and expected user load of the lip reading application.