# UNIVERSITY INSTITUTE OF COMPUTING

**DIVISION- MCA/BCA/BSc(CS)**

***Submitted by***

***SAHIL SAIFI and CHAUHAN NARESH DASU***

IN

MASTER OF COMPUTER APPLICATION

**Chandigarh University**

AUGUST 2024

# UNIVERSITY INSTITUTE OF COMPUTING

**DIVISION- MCA/BCA/BSc(CS)**

## BONAFIDE CERTIFICATE

Certified that this project report "LIP READING APPLICATION**"** is the bonafide work of **"Sahil Saifi and Chauhan Naresh Dasu**" who carried out the project work under my/our supervision.

SIGNATURE

**Mrs. Saranjeet Kaur**

SUPERVISOR

SIGNATURE

HEAD OF THE DEPARTMENT

Submitted for the project viva-voce examination held on _____

**INTERNAL EXAMINER**

**EXTERNAL EXAMINER**

# INDEX

**CONTENT** **Page no.**



## Introduction to the Project

Human communication is predominantly verbal, yet non-verbal cues such as lip movements play a significant role in understanding spoken language. Lip reading, the process of interpreting speech by visually analyzing the movements of the lips, has been a crucial skill for individuals with hearing impairments. With advancements in technology, automating lip reading through machine learning has become an emerging field of research. This project aims to develop a Lip Reading app using Tensorflow and Streamlit, combining state-of-the-art deep learning techniques with a user-friendly web interface.

Traditional lip reading relies heavily on human expertise, which can be subjective and prone to errors. Automated lip reading systems can offer a more consistent and scalable solution. However, developing an effective automated lip reading system presents several challenges:

1. **Variability in Lip Movements:** Different individuals have unique ways of articulating words, making it difficult to create a one-size-fits-all model.
2. **Environmental Factors:** Variations in lighting, camera angles, and occlusions can affect the accuracy of lip reading.
3. **Complexity of Visual Speech Recognition:** Lip reading involves understanding subtle movements and synchronizing them with spoken language, requiring sophisticated algorithms and significant computational power.

# Client Identification and Recognition of Need

## Client Identification

The development of the Lip Reading app targets a broad spectrum of clients who would significantly benefit from the capabilities of automated lip reading technology. These clients can be categorized into several key groups:

1. **Individuals with Hearing Impairments:**
- **Primary Clients:** The primary beneficiaries of this technology are individuals with hearing impairments who rely on visual cues to understand spoken language. Traditional lip reading can be challenging and inconsistent, making automated lip reading a valuable tool to enhance their communication abilities.
- **Special Education Institutions:** Schools and institutions that provide education and support to individuals with hearing impairments can integrate this technology into their teaching methods, offering students an additional resource for learning and communication.

2. **Healthcare Professionals:**
- **Audiologists and Speech Therapists:** Professionals who work with individuals with speech and hearing impairments can use the Lip Reading app as an assessment and training tool, providing real-time feedback and analysis to improve their patients' lip reading skills.
- **Clinicians and Researchers:** Researchers in the fields of speech pathology and audiology can utilize the app for studying and analyzing lip movements, contributing to advancements in their respective fields.

3. **Security and Surveillance Agencies:**
- **Law Enforcement and Security Agencies:** Automated lip reading can be employed in surveillance systems to interpret conversations in situations where audio is unavailable or of poor quality, enhancing security measures and investigative capabilities.

4. **Media and Entertainment Industry:**
- **Content Creators and Producers:** The app can be used to create more accessible content by generating accurate subtitles and captions based on lip movements, ensuring inclusivity for viewers with hearing impairments.
- **Video Conferencing Platforms:** Integration with video conferencing tools can enhance the user experience by providing real-time transcription and translation of spoken language through lip reading.

5. **Educational Institutions:**

- **Universities and Research Institutions:** Academics and researchers focusing on artificial intelligence, machine learning, and human-computer interaction can use the Lip Reading app as a case study for developing and testing advanced algorithms.

**Recognition of Need**

The need for an automated Lip Reading app arises from several critical challenges and gaps in current solutions:

1. **Inconsistent and Subjective Human Lip Reading:** Human lip reading is inherently subjective and can vary significantly based on the skill and experience of the individual. Automated lip reading offers a standardized approach, reducing inconsistencies and improving accuracy.

2. **Accessibility for Individuals with Hearing Impairments:** Communication barriers for individuals with hearing impairments can lead to social isolation and reduced opportunities in education and employment. The Lip Reading app addresses this by providing a reliable tool to understand spoken language through visual cues, promoting inclusivity and equal opportunities.

3. **Enhancement of Communication Technologies:** With the rise of remote work and virtual communication, there is an increasing demand for technologies that can enhance the clarity and accessibility of communication. Integrating automated lip reading into video conferencing platforms and multimedia content can significantly improve the user experience.

4. **Advancements in AI and Machine Learning:** The field of visual speech recognition is still in its nascent stages, with considerable potential for growth and innovation. The development of the Lip Reading app contributes to this field by exploring advanced deep learning techniques and their practical applications.

5. **Support for Healthcare and Education:** In healthcare and educational settings, automated lip reading can serve as a valuable tool for assessment, training, and research. It provides real-time feedback and analysis, aiding professionals in their work and enhancing the learning experience for students.

6. **Security and Surveillance Enhancements:** In security and surveillance, the ability to interpret lip movements can provide crucial information in situations where audio data is compromised. This capability enhances the effectiveness of security measures and investigative processes.

## Fundamental research

Fundamental research for the development of the Lip Reading app involves exploring the theoretical and practical aspects of visual speech recognition, deep learning, and the integration of these technologies

into a user-friendly application. This research encompasses various domains, including computer vision, machine learning, human-computer interaction, and accessibility. The aim is to build a solid foundation of knowledge that informs the design, development, and optimization of the Lip Reading app.

**Key Areas of Research**

1. **Visual Speech Recognition:**

- **Historical Context:** Researching the evolution of visual speech recognition, from early attempts at manual lip reading to the development of automated systems. This includes studying previous methodologies, challenges faced, and milestones achieved in the field.
- **Lip Movement Analysis:** Understanding the intricacies of lip movements during speech. This involves studying the physiology of speech production, phonetics, and the visual patterns associated with different phonemes and words.
- **Challenges in Lip Reading:** Identifying the primary challenges in visual speech recognition, such as variability in lip movements, speaker dependency, and environmental factors like lighting and occlusion. This research helps in designing robust models that can handle these challenges effectively.

2. **Deep Learning Techniques:**

- **Convolutional Neural Networks (CNNs):** Exploring the application of CNNs for spatial feature extraction from video frames. This includes studying various CNN architectures, their strengths and weaknesses, and their suitability for lip reading tasks.
- **Recurrent Neural Networks (RNNs):** Investigating the use of RNNs, particularly Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), for temporal sequence modeling. This research focuses on how these networks can capture the temporal dependencies in lip movement sequences.
- **Hybrid Models:** Examining the combination of CNNs and RNNs (e.g., CNN-RNN hybrids) to leverage both spatial and temporal features for improved performance in lip reading.
- **Transfer Learning and Pretrained Models:** Researching the use of transfer learning and pretrained models, such as those trained on large-scale video datasets, to enhance the performance and training efficiency of the lip reading model.

3. **Data Collection and Preprocessing:**

- **Dataset Compilation:** Identifying and compiling publicly available lip reading datasets. This includes datasets like GRID, TCD-TIMIT, and others that provide annotated video sequences of lip movements.
- **Data Augmentation:** Exploring techniques for augmenting the data to increase the diversity and robustness of the training set. This can include methods like random cropping, flipping, rotation, and adding noise.

- **Preprocessing Techniques:** Researching preprocessing steps such as frame extraction, lip region isolation, normalization, and alignment. These steps are crucial for ensuring the quality and consistency of the input data fed into the deep learning models.

4. **Model Training and Evaluation:**

- **Training Methodologies:** Investigating different training methodologies, including supervised learning, semi-supervised learning, and unsupervised learning. This research helps in determining the most effective approach for training the lip reading model.
- **Evaluation Metrics:** Identifying appropriate evaluation metrics for assessing the performance of the lip reading model. Metrics such as accuracy, precision, recall, F1-score, and word error rate (WER) are essential for evaluating the effectiveness of the system.
- **Benchmarking:** Conducting benchmarking studies to compare the developed model against existing state-of-the-art lip reading systems. This research helps in understanding the relative performance and identifying areas for improvement.

5. **Integration with Streamlit:**

- **User Interface Design:** Researching best practices in user interface (UI) design to create an intuitive and user-friendly web application. This involves studying UI components, layout design, and interaction patterns that enhance the user experience.
- **Real-time Processing:** Investigating techniques for real-time video processing and inference. This research focuses on optimizing the model and the application to provide real-time feedback and results to the user.
- **Deployment and Scalability:** Exploring methods for deploying the app in a scalable and reliable manner. This includes researching cloud-based deployment options, load balancing, and ensuring the system can handle multiple users concurrently.

## Project Identification

### Project Description

This project aims to develop an automated Lip Reading app that leverages deep learning techniques for visual speech recognition. The application will utilize Tensorflow for model development and Streamlit for creating an interactive web interface. The app will enable users to interpret spoken language through visual analysis of lip movements, providing a valuable tool for individuals with hearing impairments, healthcare professionals, and various other stakeholders.

### Project Scope

The project involves the end-to-end development of the Lip Reading app, including data collection, model training, interface design, and performance evaluation. Key components of the project include:

1. **Data Collection and Preprocessing:**
- Gathering and curating a diverse set of video datasets that capture various lip movements and speech patterns.
- Implementing preprocessing techniques such as frame extraction, lip region isolation, and normalization to prepare the data for model training.

2. **Model Development:**
- Designing and training deep learning models using Tensorflow. This includes experimenting with different architectures like CNNs and RNNs to accurately interpret lip movements.
- Fine-tuning the models to improve accuracy and robustness, addressing challenges such as speaker variability and environmental factors.

3. **Application Development with Streamlit:**
- Creating an interactive web application using Streamlit that allows users to upload video inputs and receive real-time lip reading results.
- Ensuring the app is user-friendly, accessible, and capable of providing accurate and timely feedback.

4. **Performance Evaluation and Optimization:**
- Evaluating the model's performance using benchmark datasets and real-world scenarios.
- Continuously optimizing the model and application based on feedback and testing to enhance reliability and usability.

**Objectives**

The primary objectives of this project are:

1. **To Develop a Robust Lip Reading Model:** Utilizing advanced deep learning techniques to create a model capable of accurately interpreting lip movements across different speakers and conditions.
2. **To Create an Interactive Web Application:** Designing a user-friendly interface with Streamlit that allows users to interact with the lip reading system seamlessly.
3. **To Evaluate and Optimize the System:** Assessing the performance of the developed system through rigorous testing and making necessary improvements to ensure high accuracy and reliability.

## Task Identification

The development of the Lip Reading app involves a series of well-defined tasks that encompass the entire lifecycle of the project. These tasks are categorized into phases, ensuring a structured and systematic approach to achieving the project objectives.

**Phase 1: Planning and Research**

1. **Define Project Scope and Objectives**
   - Identify the project's goals, deliverables, and success criteria.
   - Outline the project timeline and milestones.

2. **Conduct Literature Review**
   - Research existing lip reading technologies and methodologies.
   - Identify challenges, gaps, and opportunities in the current state of the art.

3. **Gather Relevant Resources**
   - Compile a list of academic papers, articles, and resources on visual speech recognition and deep learning.
   - Identify potential datasets and tools for data collection and preprocessing.

**Phase 2: Data Collection and Preprocessing**

1. **Identify and Compile Datasets**
   - Research and select publicly available lip reading datasets.
   - Obtain necessary permissions and licenses for dataset usage.

2. **Data Preprocessing**
   - Extract frames from video datasets.
   - Isolate the lip region using image processing techniques.
   - Normalize and align the extracted frames for consistency.

**Phase 3: Model Development**

1. **Design Initial Model Architectures**
   - Develop Convolutional Neural Network (CNN) architectures for spatial feature extraction.
   - Design Recurrent Neural Network (RNN) architectures, including LSTM and GRU, for temporal sequence modeling.

2. **Train Initial Models**
   - Train CNN and RNN models using the preprocessed and augmented datasets.
   - Experiment with different hyperparameters and architectures.

**Evaluate Model Performance**

- Assess model accuracy, precision, recall, and word error rate (WER).
- Identify areas for improvement and refine model architectures.

**Phase 4: Application Development with Streamlit**

1. **Design User Interface**

- Create wireframes and design the layout of the web application.
- Ensure the interface is user-friendly, intuitive, and accessible.

2. **Develop Web Application**
   - Implement the web application using Streamlit.
   - Integrate the trained deep learning models into the application.

3. **Implement Real-Time Processing**
   - Optimize the application for real-time video processing and inference.
   - Ensure the system provides accurate and timely feedback to users.

**Phase 5: Testing and Optimization**

1. **Conduct System Testing**
   - Perform extensive testing using benchmark datasets and real-world scenarios.
   - Identify and fix bugs, performance issues, and usability problems.

2. **Optimize Models and Application**
   - Fine-tune the models for improved accuracy and robustness.
   - Optimize the web application for speed, scalability, and reliability.

3. **User Testing and Feedback**
   - Conduct user testing sessions with individuals from the target audience.
   - Gather feedback and iterate on the design and functionality of the application.

**Phase 6: Deployment and Evaluation**

1. **Deploy Application**
   - Deploy the web application on a cloud platform.
   - Ensure proper configuration and setup for accessibility and security.

2. **Monitor and Evaluate Performance**
   - Monitor the application's performance in a live environment.
   - Evaluate user feedback and system metrics to assess overall effectiveness.

3. **Conduct Impact Analysis**
   - Analyze the impact of the Lip Reading app on the target audience.
   - Identify success stories, areas for improvement, and potential for further development.

.

## Timeline of the Project (in weeks):

**Week 1-2: Planning and Research**

- Define project scope, objectives, and methodology
- Conduct literature review and gather relevant resources
- Identify and compile datasets for lip reading

**Week 2-3: Data Collection and Preprocessing**

- Collect and preprocess video datasets
- Implement data augmentation techniques
- Prepare data for model training

**Week 3-5: Model Development**

- Design and train initial deep learning models using Tensorflow
- Experiment with different architectures and hyperparameters
- Fine-tune models based on performance evaluation

**Week 5-7: Application Development**

- Develop the interactive web application using Streamlit
- Integrate trained models with the application
- Design and implement the user interface

**Week 7-8: Testing and Optimization**

- Conduct rigorous testing of the system using benchmark datasets and real-world scenarios
- Optimize models and application based on feedback and testing results
- Ensure robustness, reliability, and user-friendliness

**Week 8-9: Deployment and Evaluation**

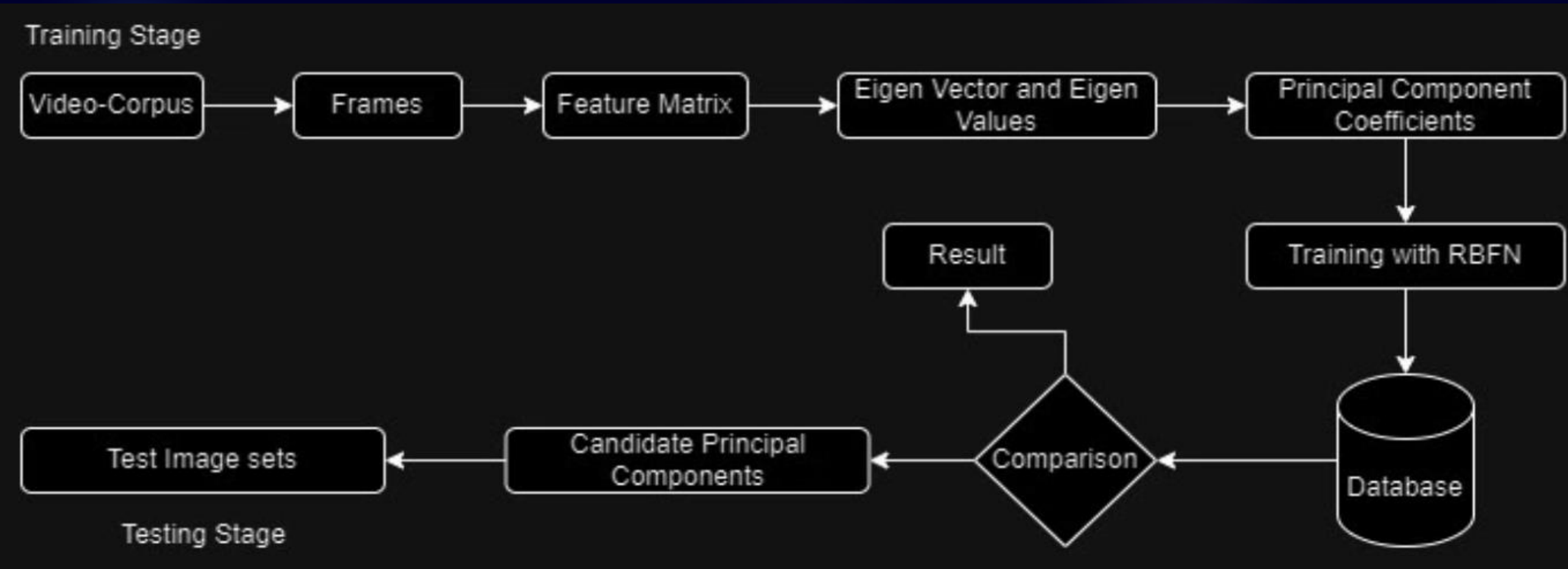
- Deploy the application on a cloud platform
- Conduct user testing and gather feedback
- Evaluate overall performance and impact of the system

# Lip Reading Application

This project explores the development of a lip reading application by Naresh and Sahil. The goal is to create a system capable of accurately transcribing spoken words based solely on lip movements.

**by Naresh Chauhan (23MCI10007) and Sahil Saifi (23MCI10142)**

# Block Diagram of the Application



# Experimental Setup and Design

1

## Dataset Collection

Gather a large dataset of lip movements associated with different words or phrases.

2

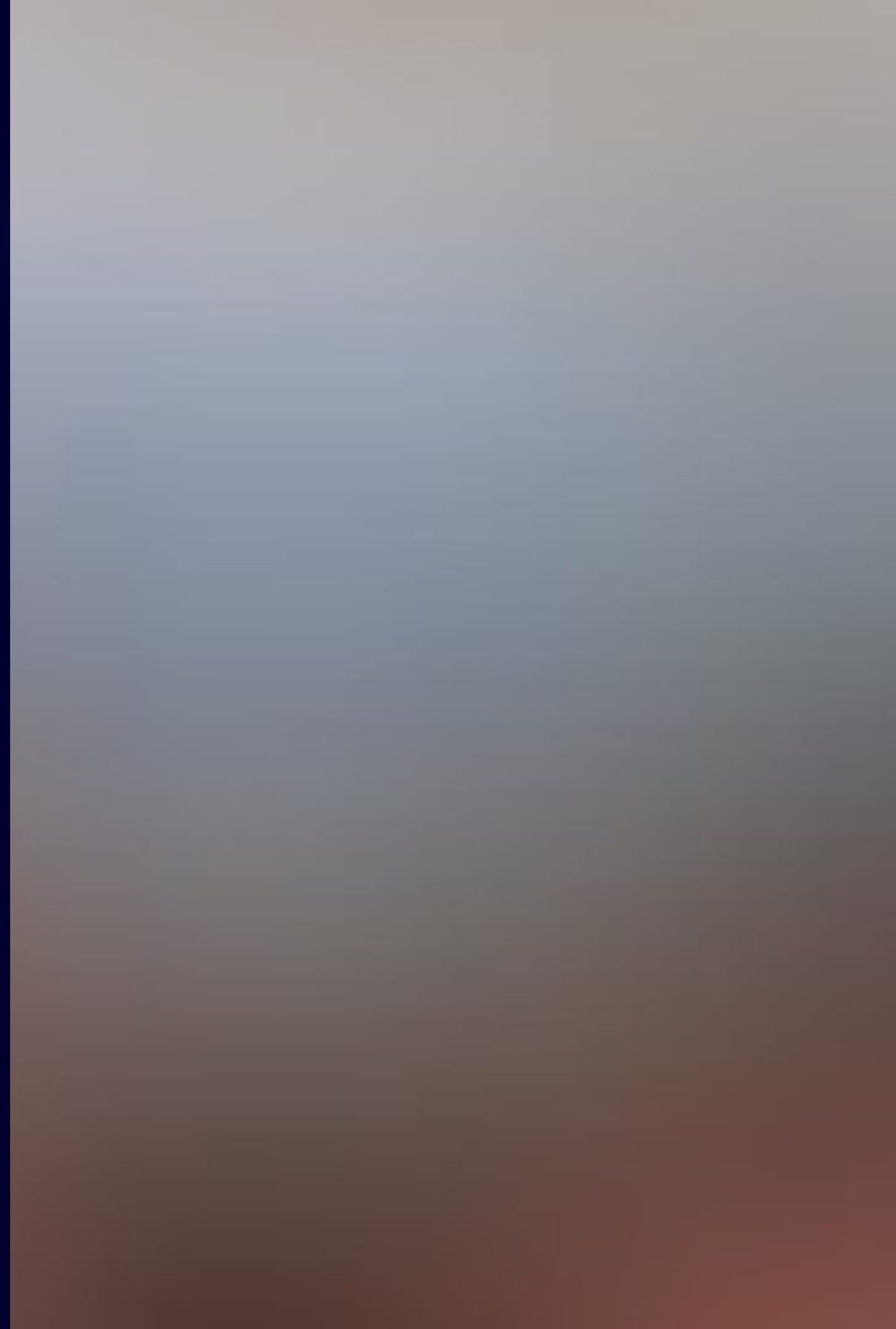
## Model Training

Train a deep learning model using the collected dataset. The model learns to associate specific lip movements with corresponding spoken words.

3

## Evaluation and Optimization

Evaluate the trained model's performance using a separate test dataset. Adjust the model parameters or the training process to improve accuracy.



# Required Equipment

1 Camera

A high-quality camera capable of capturing clear video footage of the speaker's face.

2 Computer

A powerful computer with sufficient processing power and memory to handle the computational demands of the application.

3 Microphone (Optional)

A microphone can be used to capture audio alongside the video, which can be helpful for evaluation and comparison purposes.

4 Lighting

Proper lighting is essential for the camera to capture clear and well-lit images of the speaker's face.

# Required Software

## Streamlit

A Python library for building interactive web applications that can be used to create a user-friendly interface for the lip reading application.

## TensorFlow

A popular open-source machine learning framework that provides tools for building and training deep learning models.

## OpenCV

A library for computer vision tasks such as image processing, video analysis, and object detection, which is crucial for lip segmentation and feature extraction.

## NumPy

A fundamental library for numerical computing in Python, providing powerful data structures and functions for manipulating arrays and matrices.

## ImageIO

Used for handling image and video data formats, enabling the application to load, process, and save images and videos effectively.

## Python

The primary programming language used for developing the entire application, leveraging its extensive libraries for machine learning, computer vision, and web development.

# Lip Reading Application: A Journey

This presentation outlines the development of a lip-reading application using Python, machine learning, Streamlit, and OpenCV. The goal is to create an accessible and user-friendly tool that can assist individuals with hearing impairments in understanding spoken language.

by Sahil Saifi(23MCI10142)
Naresh Chauhan(23MCI10007)

# Progress Relative to Initial Plan

1. **Initial Planning**
   The initial plan involved gathering data, selecting the appropriate machine learning model, and designing the user interface.

2. **Data Acquisition and Preprocessing**
   A substantial dataset of lip movements synchronized with corresponding audio was collected and preprocessed, ensuring high-quality data for model training.

3. **Model Selection and Training**
   A deep learning model, specifically a Convolutional Neural Network (CNN), was chosen and trained on the prepared dataset. This model was designed to recognize patterns in lip movements.

4. **User Interface Development**
   The Streamlit library was used to create a user-friendly interface, allowing users to input live video streams or upload videos for lip-reading analysis.

5. **Testing and Refinement**
   The application was rigorously tested and refined, ensuring accuracy and improving the overall performance of the lip-reading system.

# Preliminary Results or Prototype

### Accuracy

Initial tests indicate a promising level of accuracy, with the model successfully transcribing spoken words from lip movements.

### Real-Time Performance

The prototype demonstrates efficient real-time performance, capable of processing video streams and generating transcriptions without significant delays.

### User Experience

The user interface is designed for ease of use, providing a seamless experience for both inputting video data and viewing the resulting transcriptions.

# Methodology Refinement

1

## Data Augmentation

Techniques such as rotation, scaling, and color variations were incorporated to augment the training dataset, increasing model robustness.

2

## Model Optimization

The CNN model was further optimized by fine-tuning hyperparameters and exploring different architectures to enhance accuracy and performance.

3

## Error Analysis

In-depth analysis of errors helped identify areas for improvement and guide further refinements in the model's training and architecture.

4

## Integration with OpenCV

OpenCV library was integrated to provide real-time video processing and lip detection, enhancing the application's ability to extract relevant features from the input video.

# Next Steps and Future Plans

## 1 Expand Dataset

The dataset will be expanded to include a wider range of speakers, accents, and environmental conditions, further improving model generalization.

## 2 Improve Accuracy

Continuous efforts will be made to improve model accuracy by exploring advanced techniques such as transfer learning and ensemble methods.

## 3 Enhance User Interface

The user interface will be refined to incorporate additional features such as speech-to-text conversion, language translation, and accessibility enhancements.

## 4 Real-World Applications

Explore potential applications in diverse settings such as assistive technology, education, and communication systems for individuals with hearing impairments.

# Rubric 4
# (LIP READING APP)

By Naresh Chauhan Dasu (23MCI10007) and Sahil Saifi (23MCI10142)

# Agenda

1. Planned Outcomes Attainment
2. Quality and Depth of work
3. Data Analysis and Interpretation
4. Innovation and Creativity
5. Practical Applications

# PLANNED OUTCOMES AND ATTAINMENT

- **OBJECTIVE:**

Develop a system capable of accurately transcribing spoken words based solely on lip movements, improving accessibility for hearing-impaired individuals.

- **KEY STEPS:**

  - Dataset Collection: Large dataset of lip movements for various words.
  - Model Training: CNN-based model for recognizing lip patterns.
  - User Interface: Streamlit-based web app for input and output processing.

# QUALITY AND DEPTH OF WORK

## DATA COLLECTION & PREPROCESSING

- Comprehensive dataset collection synchronized with corresponding audio.
- Preprocessing includes high-quality data for efficient model training.

## MODEL TRAINING

- Deep learning (CNN) trained to recognize lip movements with real-time video processing.
- Model optimization by fine-tuning parameters, exploring architectures, and performing error analysis.

## TESTING AND REFINEMENT

- Rigorous testing for accuracy and performance.
- Refining the UI to ensure ease of use and seamless interaction.

# DATA ANALYSIS AND INTERPRETATION

- **ERROR ANALYSIS**

Identified critical areas for improvement in model training.

- **PERFORMANCE METRICS:**
  - Initial tests show promising accuracy for transcribing words based on lip movements.
  - Real-time processing with minimal delays.
- **TESTING AND REFINEMENT**

Refined using data augmentation (rotation, scaling, etc.) to increase robustness and improve error correction.

# INNOVATION AND CREATIVITY

- **TECNOLOGICAL INNOVATION:**
  - Integration of OpenCV for real-time video and lip detection.
  - Innovative use of CNN models to decode lip patterns into meaningful transcriptions.

- **CREATIVITY IN SOLUTIONS:**
  - Unique solution for hearing-impaired individuals, transforming visual lip patterns into text in real-time.
  - Exploring future integration with features like speech-to-text conversion and language translation.

# PRACTICAL APPLICATIONS

- **ASSISTIVE TECHNOLOGY:**

Direct applications for people with hearing impairments, enabling better communication.

- **BROADER USES:**
  - Potential use cases in education, communication, and accessibility systems.
  - Real-world utility in environments where audio isn't available or viable.

- **FUTURE EXPANSION:**

Planned expansion of datasets to include various accents, environments, and further improvements in accuracy.

# THANK YOU!