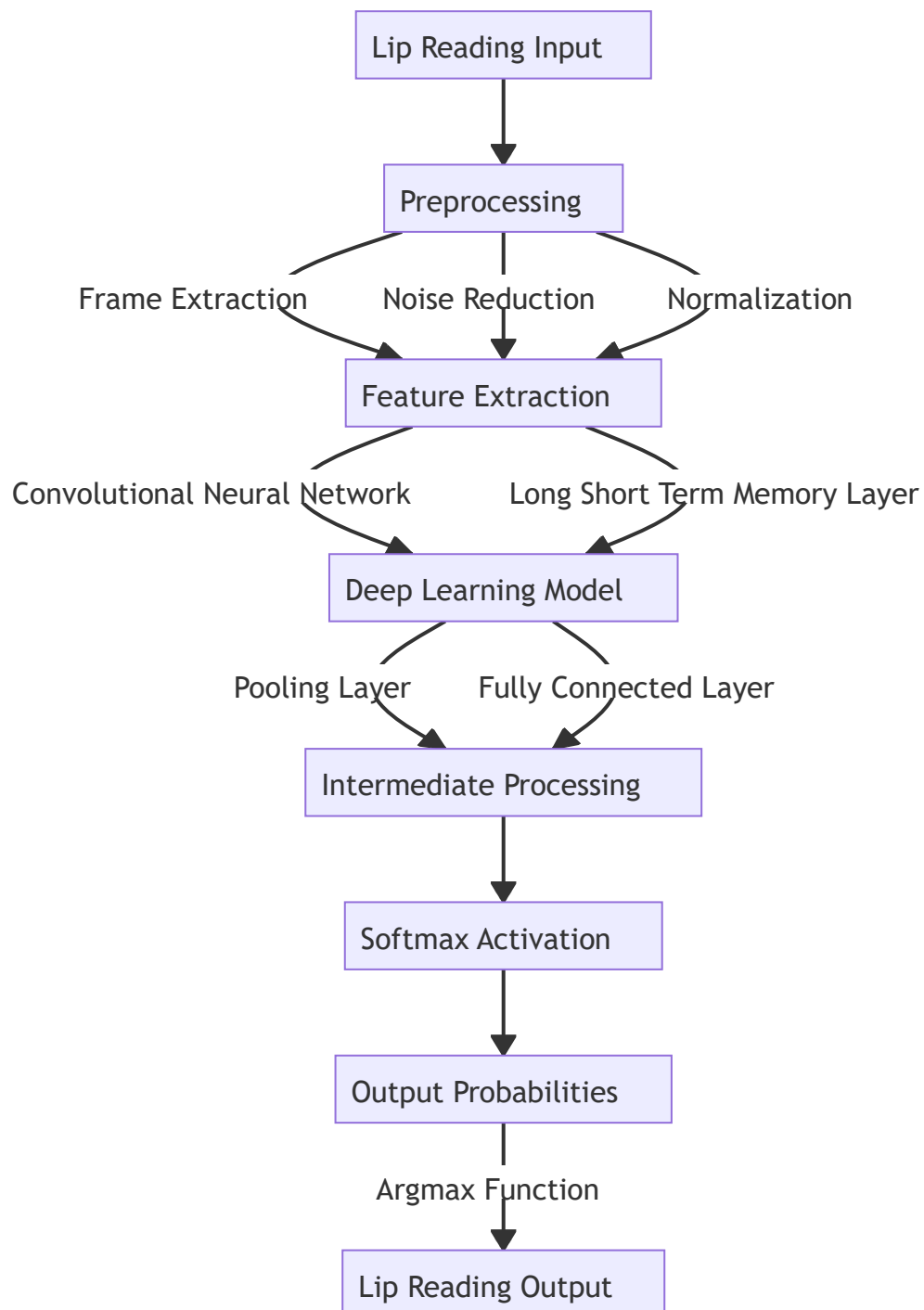


Project Design Phase-I
Solution Architecture

Date	23 October 2023
Team ID	Team-592895
Project Name	Lip reading using Deep Learning
Maximum Marks	5 Marks

Solution Architecture



1. Lip Reading Input: This section represents the source of data for the lip-reading system. It typically includes video or image frames containing a person's lips, which serve as the input for the lip-reading process.
2. Preprocessing: Preprocessing involves the initial data cleaning and enhancement steps. This can include tasks like noise reduction, image stabilization, and frame alignment to prepare the lip data for subsequent analysis.
3. Feature Extraction: In this stage, relevant features are extracted from the pre-processed lip data. These features might include lip shape, motion patterns, or other visual cues that are essential for lip reading.
4. Deep Learning Model: This section represents the core of the architecture. It outlines the neural network model used for lip reading. This model is typically a deep learning architecture, such as a convolutional neural network (CNN) or recurrent neural network (RNN), designed to learn and understand the extracted features.
5. Intermediate Processing: This stage refers to any additional data transformations or computations that occur within the deep learning model. It might include hidden layers, recurrent connections, or other processing steps within the neural network.
6. Softmax Activation: The softmax activation is often applied at the output layer of the neural network. It converts the model's raw predictions into probability distributions over different classes or phonemes, which represent the likelihood of each phoneme being spoken based on the lip movements.
7. Output Probabilities: This part shows the probabilities generated by the softmax activation layer. Each probability corresponds to the likelihood of a specific phoneme or word being spoken based on the lip movements observed in the input data.
8. Lip Reading Output: This is the final output of the lip-reading system. It typically involves selecting the phoneme or word with the highest probability from the softmax output as the recognized lip-reading result. The output can be in the form of text or phonetic transcription, representing what was spoken by the person based on their lip movements.

This architecture chart helps visualize the sequential stages involved in the lip-reading process, from data input to the final lip reading output, highlighting the role of deep learning and neural networks in deciphering spoken words from lip movements.