

## **ABSTRACT**

Mahesa Tirta Panjalu, 51420345

### ***LIP READING USING CNN, BI-LSTM AND CTC LOSS FUNCTION ON ENGLISH DATASET***

*Thesis, Department of Informatics, Faculty of Industrial Technology, Gunadarma University, 2024*

**Keywords:** *Lip Reading, LipNet, CNN, LSTM, CTC, WER, CER*

*(xiv + 77 Pages + Attachment)*

*Deaf people are people who have hearing loss. The main impact of this condition is a barrier in verbal or oral communication, making it difficult to communicate with hearing people. The lips are the part commonly used to speak or communicate. Lip movements when communicating will produce different movements for each word or letter spoken. Lips can be used to predict words from lip movements that will be detected when speaking. Increasingly developing technology can help these problems in reading lips. Convolutional Neural Network or CNN has grown rapidly and become one of the most popular methods in the field of image recognition and video processing due to its ability to automatically learn features from input data. This research aims to perform lip reading using CNN, Long Short-Term Memory (LSTM) and Connectionist Temporal Classification (CTC) methods in English. This research uses a dataset from The Grid audiovisual sentence corpus of 1000 videos and 1000 texts. The preprocessing stage consists of two parts, namely video preprocessing and text preprocessing. The video preprocessing stage includes grayscale conversion, frame cropping, augmentation and normalization. The text preprocessing stage carried out the encoding process on the alignments dataset. The classification stage uses Convolutional Neural Networks, Long Short-Term Memory and Connectionist Temporal Classification Loss Function methods. The evaluation results get the best accuracy value of 96.9%, Word Error Rate (WER) of 0.66%, and Character Error Rate (CER) of 0.16% using a model with 80:20 data scenario and batch size 2.*

References (2016-2024)