

DEEP LEARNING BASED MULTI-MODAL EMOTION VARIATION DETECTION

Rubasin Siriwardhana Vinuri Thisara

A dissertation submitted in partial fulfilment of the requirement for
Bachelor of Science (Honours) degree in Computer Science

School of Computing

**Informatics Institute of Technology, Sri Lanka in collaboration with
University of Westminster, UK**

2022

Abstract

People need to be able to talk about their feelings in a healthy way, especially in healthcare and counselling settings. But because of things like social rules, anxiety, and cultural background, people often find it hard to show how they really feel. To deal with this problem, we have come up with a Multi-Modal Emotion Recognition System that uses video analysis to instantly find and analyses how people are feeling.

The first thing the system does is check the type of video file and make sure there is sound and faces in the video. If these conditions are met, the video goes through preprocessing steps that include removing frames and audio clips and reshaping and resizing them to fit the model. The model consists of a Convolutional Neural Network (CNN) that leverages deep learning techniques to learn and extract features from the visual and audio signals of the video.

After the processing steps, the system shows the changes in the person's emotions in real time and makes a graph of how those changes have changed over time. The system also gives an overall report of the feelings shown in the video, including the first emotion, the most common emotion throughout the video, and the last emotion. This knowledge can help therapists and psychiatrists figure out how their patients are feeling without making them feel uncomfortable.

The system is evaluated on a dataset of videos that were annotated by human experts, and achieved high accuracy in detecting various emotions such as happiness, sadness, anger, and surprise. The suggested system can be used in places like mental health clinics, schools, and law enforcement to make diagnosis, counselling, and investigations more accurate. But when recording and analyzing people's behaviors, social questions about privacy and permission must be carefully thought through and dealt with.

Keywords : Deep Learning ,CNN, Multi-Modal, Emotion Recognition, Emotion Variation, Speech Emotion variation, Video Emotion Recognition, System for Health