CHAPTER I

INTRODUCTION

1.1 Research Background

Facial Expression are among the most basic and common types of non-verbal communication in people. They inherently indicate an individual's feelings, motives, and mental reactions. (Ekman & Friesen, 1978).

The capability to instantly identify and analyze facial expressions has opened up numerous groundbreakings uses in different industries. Within human-computer interaction, technologies that grasp user feelings can facilitate more tailored and responsive experiences, such as in virtual help systems or social robots. In the realm of security, noticing facial expressions can definitely serve as an initial sign of potentially suspicious actions. In the educational sector, observing the facial expressions of students can assist educators in assessing their understanding and involvement levels. (S et al., 2022.).

Even with its wide range of possible uses, the automatic recognition of facial expressions encounters notable difficulties. The complexities in this field stem from natural differences in lighting, angles from which faces are seen, obstructions, and the unique ways people show their feelings. Conventional techniques for recognizing facial expressions typically depend on extracting features by hand and using simple machine learning, which usually struggles with this kind of variability and demands considerable specialized knowledge.

Rapid advances in deep learning, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable performance in computer vision tasks, including image classification. CNNs are highly effective in extracting hierarchical features from image data. But Facial expressions are fluid, consisting of a series of changes in facial characteristics as time progresses. To effectively recognize this time-related component, a model that can handle sequential information is necessary. (Dewi & Ismawan, 2021).

Long Short-Term Memory (LSTM), as a type of Recurrent Neural Network (RNN), is well-suited for modeling long-term dependencies in sequential data. The CNN-LSTM combination, potentially combining the spatial feature extraction capabilities of CNN with the sequential modeling capabilities of LSTM, is expected to improve the accuracy of facial expression classification, especially on datasets that capture spatial and temporal variations (Altiarika & Sari, 2023).

The swift advancement of technology in artificial intelligence, especially in deep learning, has created new opportunities for analyzing human behavior. Deep learning's capacity to identify intricate patterns from extensive raw data, without requiring manually created features, positions it as a perfect choice for detecting facial expressions. This research seeks to evaluate how well a standard CNN model performs compared to a CNN-LSTM model in classifying facial expressions using the FER2013 dataset. Consequently, it aims to shed light on how effective each architecture is in managing the challenges associated with classifying facial expressions.

1. 2 Problem Identification

Based on the above background, this study identifies several key issues. While CNN models are highly effective in extracting spatial features from images, they are limited in capturing the temporal information inherent in dynamic facial expressions. Therefore, it is crucial to understand how the integration of sequential architectures such as LSTMs can enhance spatial features. Furthermore, a clear comparative evaluation between CNNs and CNN-LSTMs on standard datasets such as FER2013 is needed to establish a benchmark, given the challenges in distinguishing facial expressions with visual similarity or low intensity.

By addressing these gaps, this research aims to provide a more robust framework for real-time facial expression recognition. Ultimately, bridging the gap between spatial and temporal feature extraction will not only improve classification accuracy but also offer deeper insights into deploying these hybrid models for practical, real-world applications.

1. 3 Research Limitation

In order for the research to be more focused, several limitations are applied, namely:

a. The dataset used is FER2013, which contains grayscale facial images and has been labeled in seven basic expression categories (anger, disgust, fear, happiness, sadness, surprise, neutral).
b. The methods to be compared are limited to CNN and CNN-LSTM.
c. Classification will be done only based on facial image data.

1. 4 Research Question

Based on the background above, the problem formulation in this research is:

a. How does the CNN model perform in classifying 7 basic facial expressions based on the FER2013 Dataset?

b. How does the CNN-LSTM model perform in classifying 7 basic facial expressions based on the FER2013 Dataset?
c. How does the performance of CNN and CNN-LSTM models compare in classifying 7 basic facial expressions based on the FER2013 Dataset, seen from the metrics of accuracy, precision, recall, F1-score, and AUC?

1. 5 Research Objectives

The objectives of this research are:

a. Implementing and evaluating the performance of a CNN model in classifying 7 basic facial expressions using the FER2013 Dataset.
b. Implementing and evaluating the performance of the CNN-LSTM model in classifying 7 basic facial expressions using the FER2013 Dataset.
c. Conducting a performance comparison between CNN and CNN-LSTM models in classifying 7 basic facial expressions based on the FER2013 Dataset.

1. 6 Research Significances

This research is expected to provide the following benefits:

a. Theoretical Benefits: Adds to the body of research regarding deep learning uses in facial expression recognition, particularly through a comparison of CNN and CNN-LSTM architectures. The findings from this research can provide a foundation for creating advanced models down the line.
b. Practical Benefits: The performance comparison results can guide researchers and developers in selecting an appropriate deep learning architecture for facial expression classification tasks, especially when

considering temporal aspects. This can be applied in the development of human-computer interaction systems, sentiment analysis, or security applications.

c. Methodological Benefits: explaining in detail the stages and methods used, so that they can serve as a reference for similar research in the future.