IMPLEMENTASI METODE *CONVOLUTIONAL NEURAL NETWORK* UNTUK PENGENALAN EMOSI BERBASIS SUARA MANUSIA MENGGUNAKAN REPRESENTASI CITRA *SPECTROGRAM*

ENDOW BONAPEN

ABSTRAK

Pengenalan emosi berdasarkan suara (*Speech Emotions Recognition)*, merupakan pengenalan emosi berdasarkan suara manusia yang dihasilkan dari keadaan atau situasi yang tengah dialaminya. Dalam kehidupan sehari-hari, ragam emosi yang diungkapkan melalui suara dapat berbeda antara individu satu dan lainnya. Namun, bagaimana komputer dapat memiliki kapabilitas untuk membedakan ragam emosi yang terkandung dalam sinyal suara seperti manusia yang dapat membedakannya. Dari permasalahan tersebut, penelitian ini bertujuan untuk membuat dan membangun model pengenalan emosi berdasarkan suara manusia, model tersebut dapat digunakan untuk membedakan dan mengetahui ragam emosi manusia. Maka dari itu, dalam penelitian ini akan menerapkan fitur ekstraksi *Mel Frequency Cepstral Coefficients* (MFCC) *Spectrogram*, kemudian hasil dari fitur ekstraksi akan masuk kedalam proses pengenalan emosi dengan menerapkan *Deep Learning*, metode yang digunakan adalah *Convolutional Neural Network* (CNN) 2D. Jenis ragam emosi yang digunakan dalam penelitian ini adalah *Angry, Disgust, Fear, Happy, Sad, Neutral,* dan *Surprise*. Pada penelitian ini dilakukan proses pengumpulan data dengan total data 4665 data suara, *pre-process* data suara, pembagian rasio data, proses pelatihan dengan menggunakan data latih dan validasi, dan proses pengujian dengan menggunakan data uji. Hasil dalam membangun model pengenalan emosi berdasarkan suara manusia dengan menerapkan metode CNN menghasilkan akurasi tertinggi mencapai 88%, dengan nilai *precision* 90%, *recall* 87%, dan *f1 – score* 88%. Akurasi pada setiap jenis ragam emosi yaitu *Angry* 96,80%, *Disgust* 97,86%, *Fear* 96,16%, *Happy* 95,52%, *Neutral* 96, 37%, *Sad* 96,80%, *Surprise* 95,52%.

Kata Kunci: *Speech Emotion Recognition*, *Spectrogram*, CNN 2D, MFCC.

IMPLEMENTATION OF CONVOLUTIONAL NEURAL NETWORK METHOD FOR HUMAN VOICE-BASED EMOTION RECOGNITION USING SPECTROGRAM IMAGE REPRESENTATION

ENDOW BONAPEN

ABSTRACT

Speech Emotion Recognition (SER) is the identification of human emotions based on the voice generated during a particular state or situation. In everyday life, the range of emotions expressed through voice can vary among individuals. However, how can a computer have the capability to distinguish the variety of emotions contained in a sound signal, much like humans can discern them? Addressing this issue, this research aims to create and build a model for recognizing human emotions based on voice, which can be used to differentiate and identify various human emotions. Therefore, this study will apply Mel Frequency Cepstral Coefficients (MFCC) Spectrogram as a feature extraction method, and the extracted features will be processed through emotion recognition using Deep Learning. The chosen method is the 2D Convolutional Neural Network (CNN). The emotions considered in this research include Angry, Disgust, Fear, Happy, Sad, Neutral, and Surprise. The data collection process involves a total of 4665 voice samples, followed by pre-processing, data ratio splitting, training using training and validation data, and testing using test data. The results of building the emotion recognition model based on human voice using the CNN method achieved the highest accuracy of 88%, with precision at 90%, recall at 87%, and an F1-score of 88%. The accuracy for each emotion type is as follows: Angry 96.80%, Disgust 97.86%, Fear 96.16%, Happy 95.52%, Neutral 96.37%, Sad 96.80%, and Surprise 95.52%.

Keywords: *Speech Emotion Recognition*, *Spectrogram*, CNN 2D, MFCC.