

**PENGENALAN EMOSI SUARA MANUSIA MENGGUNAKAN
CONVOLUTIONAL NEURAL NETWORK DENGAN KOMBINASI
EKSTRAKSI FITUR MFCC DAN GFCC**

GUNTUR LAKSONO PUTRA

ABSTRAK

Emosi merupakan salah satu faktor penting dalam interaksi antar manusia, sering kali emosi mempengaruhi tingkah laku manusia. Emosi dapat dikenali melalui ekspresi wajah, karakteristik suara, dan bahasa tubuh seseorang. Seiring dengan perkembangan teknologi, manusia terus berusaha melakukan penelitian agar komputer dapat memahami keinginan dan perasaan manusia. Salah satu teknologi yang sedang dikembangkan adalah kemampuan komputer untuk mengenali emosi melalui informasi yang terkandung dalam suara manusia. Penelitian ini mencoba mengimplementasikan metode CNN dengan ekstraksi fitur kombinasi MFCC dan GFCC untuk mengklasifikasikan emosi dari pembicara menggunakan *dataset Toronto Emotional Speech Set* yang terdiri dari 7 kelas emosi. Hasil penelitian menunjukkan bahwa penggunaan kombinasi fitur MFCC+GFCC secara positif mempengaruhi kinerja model klasifikasi CNN, dengan mencapai nilai rata-rata akurasi sebesar 98.57%, rata-rata presisi sebesar 98.65%, dan rata-rata *recall* sebesar 98.57%. Penggunaan kombinasi fitur MFCC+GFCC juga meningkatkan rata-rata akurasi dan *recall* sebesar 2.62%, serta meningkatkan rata-rata presisi sebesar 2.35-2.47% dibandingkan dengan penggunaan fitur tunggal. Temuan ini menunjukkan potensi penggunaan kombinasi fitur MFCC+GFCC dalam pengenalan emosi suara manusia, dengan implikasi penting dalam pengembangan sistem pengenalan suara yang lebih canggih dan responsif terhadap emosi manusia.

Kata kunci: Pengenalan Emosi; Suara Manusia; MFCC; GFCC; CNN

HUMAN VOICE EMOTION RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK WITH A COMBINATION OF MFCC AND GFCC FEATURE EXTRACTION

GUNTUR LAKSONO PUTRA

ABSTRACT

Emotions are an important factor in interactions between humans, emotions often influence human behavior. Emotions can be recognized through a person's facial expressions, voice characteristics, and body language. Along with the development of technology, humans continue to try to conduct research so that computers can understand human desires and feelings. One technology that is being developed is the ability of computers to recognize emotions through the information contained in the human voice. This research tries to implement the CNN method with combined MFCC and GFCC feature extraction to classify speakers' emotions using the Toronto Emotional Speech Set dataset which consists of 7 emotion classes. The research results show that the use of the MFCC+GFCC feature combination positively influences the performance of the CNN classification model, by achieving an average accuracy value of 98.57%, an average precision of 98.65%, and an average recall of 98.57%. Using a combination of MFCC+GFCC features also increases the average accuracy and recall by 2.62% and increases the average precision by 2.35-2.47% compared to using a single feature. These findings demonstrate the potential of using the combination of MFCC+GFCC features in human voice emotion recognition, with important implications in the development of more sophisticated speech recognition systems responsive to human emotions.

Keywords: *Emotion Recognition; Human Voice; MFCC; GFCC; CNN*