

SKRIPSI



**Pengenalan Emosi Suara Manusia Menggunakan
Convolutional Neural Network Dengan Kombinasi
Ekstraksi Fitur MFCC dan GFCC**

GUNTUR LAKSONO PUTRA

2010511033

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA

2024

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Komputer Pada Fakultas Ilmu Komputer**



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

GUNTUR LAKSONO PUTRA

2010511033

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA

2024

PERNYATAAN ORISINALITAS

Tugas akhir ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Guntur Laksono Putra

NIM : 2010511033

Tanggal : 27 Mei 2024

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 27 Mei 2024

Yang menyatakan,



(Guntur Laksono Putra)

PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai civitas akademik Universitas Pembangunan Nasional “Veteran” Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Guntur Laksono Putra

NIM : 2010511033

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional “Veteran” Jakarta Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul :

Pengenalan Emosi Suara Manusia Menggunakan Convolutional Neural Network dengan Kombinasi Ekstraksi Fitur MFCC dan GFCC

Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian Pernyataan ini saya buat dengan sebenarnya.

Jakarta, 27 Mei 2024

Yang menyatakan,



(Guntur Laksono Putra)

LEMBAR PENGESAHAN

Skripsi diajukan oleh:

Nama : Guntur Laksono Putra

NIM : 2010511033

Program Studi : S1 Informatika

Judul Skripsi : Pengenalan Emosi Suara Manusia Menggunakan Convolutional Neural Network Dengan Kombinasi Ekstraksi Fitur MFCC dan GFCC

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Ridwan Raafi'udin, S.Kom., M.Kom.

Penguji 1



Iin Ernawati, S.Kom., M.Si.

Penguji 2



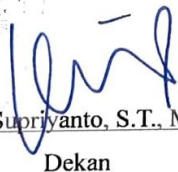
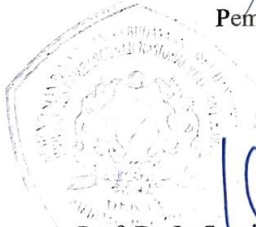
Jayanta, S.Kom., M.Si.

Pembimbing 1



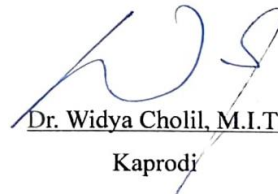
Nurul Afifah Arifuddin, S.Pd., M.T.

Pembimbing 2



Prof. Dr. Ir. Suwiyanto, S.T., M.Sc., IPM

Dekan



Dr. Widya Cholil, M.I.T.

Kaprodi

Ditetapkan di : Jakarta

Tanggal Ujian : 24 April 2024

**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*

KATA PENGANTAR

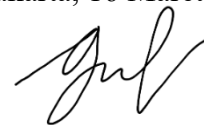
Segala puji dan syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan nikmat sehat dan nikmat panjang umur, sehingga penulis dapat menyelesaikan pengerjaan proposal skripsi ini yang berjudul “Pengenalan Emosi Suara Manusia Menggunakan *Convolutional Neural Network* Dengan Kombinasi Ekstraksi Fitur MFCC dan GFCC”. Penulisan skripsi ini bertujuan untuk mendalami dan memperluas pengetahuan mengenai pengenalan emosi suara manusia, sebuah bidang yang memiliki implikasi signifikan dalam berbagai aplikasi teknologi, terutama dalam konteks teknologi pengenalan suara.

Penulis menyadari bahwa penulisan skripsi ini tidak luput dari segala bentuk bantuan, dukungan, dan bimbingan dari berbagai pihak, sehingga penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah membantu dan membimbing penulis selama pengerjaan skripsi ini, terutama kepada:

1. Pak Jayanta S.Kom., M.Si., selaku dosen pembimbing I yang telah memberikan saran dan arahan selama penyusunan skripsi ini.
2. Ibu Nurul Afifah Arifuddin S.Pd., M.T, selaku dosen pembimbing II yang telah mengoreksi penulisan pada skripsi ini.
3. Keluarga, yang senantiasa memberikan dukungan baik moral dan materil untuk kelancaran penyusunan skripsi ini.
4. Rekan seperjuangan, yang telah memberikan banyak motivasi dan inspirasi untuk menyelesaikan penyusunan skripsi ini.

Penulis berharap skripsi ini dapat memberikan dampak positif dan manfaat dalam pengembangan ilmu pengetahuan khususnya dalam bidang pengenalan emosi suara manusia. Selain itu, penulis ingin menyampaikan permohonan maaf apabila dalam penulisan skripsi ini terdapat kesalahan yang tidak penulis sadari.

Jakarta, 16 Maret 2024



Guntur Laksono Putra

DAFTAR ISI

LEMBAR COVER	i
LEMBAR JUDUL	ii
PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI	iv
LEMBAR PENGESAHAN	v
ABSTRAK	vi
<i>ABSTRACT</i>	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Luaran Yang Diharapkan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Emosi	5
2.2 Suara	5
2.3 Ekstraksi Fitur	6
2.3.1 <i>Mel Frequency Cepstral Coefficient</i> (MFCC)	7
2.3.2 <i>Gammatone Frequency Cepstral Coefficient</i> (GFCC)	11
2.4 <i>Convolutional Neural Network</i> (CNN)	13
2.4.1 <i>Convolutional Layer</i>	14
2.4.2 <i>Activation Function</i>	15
2.4.3 <i>Pooling Layer</i>	15

2.4.4	<i>Flattening Layer</i>	16
2.4.5	<i>Fully Connected Layer</i>	16
2.5	<i>Confusion Matrix</i>	17
2.6	Penelitian Terdahulu	18
BAB III METODE PENELITIAN		21
3.1	Kerangka Berpikir	21
3.1.1	Pengumpulan Data	22
3.1.2	Ekstraksi Fitur	22
3.1.3	Klasifikasi Emosi	24
3.1.4	Evaluasi Model	25
3.2	Perangkat Penelitian	25
3.2.1	Perangkat Keras	26
3.2.2	Perangkat Lunak	26
3.3	Jadwal Penelitian	26
BAB IV HASIL DAN PEMBAHASAN		27
4.1	Pengumpulan Data	27
4.2	Ekstraksi Fitur	29
4.3	Kalsifikasi Emosi	44
4.4	Evaluasi Model	50
BAB V SIMPULAN DAN SARAN		55
5.1	Kesimpulan	55
5.2	Saran	56
DAFTAR PUSTAKA		57

DAFTAR GAMBAR

Gambar 2.1. Roda emosi Plutchik (Plutchik, 2001).....	5
Gambar 2.2. (a) Sinyal analog dan (b) sinyal digital	6
Gambar 2.3. Diagram alir MFCC	7
Gambar 2.4. Ilustrasi proses <i>framing</i> (Santoso et al., 2022).....	8
Gambar 2.5. <i>Hamming window</i> (Santoso et al., 2022).....	9
Gambar 2.6. <i>Mel filter bank</i> (Fayek, 2016).....	10
Gambar 2.7. Diagram alir GFCC.	12
Gambar 2.8. <i>Gammatone filter bank</i> (Lian et al., 2017)	13
Gambar 2.9. Arsitektur CNN 1D (Sánchez-Reolid et al., 2022).....	14
Gambar 2.10. Proses konvolusi.....	15
Gambar 2.11. Operasi <i>max pooling</i>	16
Gambar 3.1. Kerangka berpikir.....	21
Gambar 3.2. Arsitektur model CNN	24
Gambar 4.1. Grafik distribusi data suara berdasarkan emosi.....	27
Gambar 4.2. <i>Raw waveform</i> dan <i>pre-emphasised waveform</i>	30
Gambar 4.3. Ilustrasi proses <i>framing</i> pada sampel data berdurasi 60 ms.	31
Gambar 4.4. Sampel tiga <i>frame</i> hasil proses <i>framing</i>	31
Gambar 4.5. <i>Hamming window</i> seukuran <i>frame</i>	32
Gambar 4.6. Ilustrasi proses <i>windowing</i>	32
Gambar 4.7. Sampel tiga <i>frame</i> hasil proses <i>windowing</i>	32
Gambar 4.8. Hasil proses FFT dengan simetris (a) dan tanpa simetrisnya (b)	33
Gambar 4.9. Tiga <i>frame</i> berurut hasil proses FFT tanpa sisi simetrisnya.	34
Gambar 4.10. <i>Mel filter bank</i> (40 filter).....	36
Gambar 4.11. Ilustrasi proses <i>filtering</i> dengan salah satu filter <i>mel filter bank</i>	37

Gambar 4.12. Sampel energi hasil proses <i>mel filter bank</i>	37
Gambar 4.13. <i>Gammatone filter bank</i> (40 filter).	38
Gambar 4.14. Proses <i>filtering</i> dengan <i>filter gammatone</i>	38
Gambar 4.15. Sampel energi hasil proses <i>gammatone filter bank</i>	38
Gambar 4.16. <i>Log scaled mel filter bank energy</i>	39
Gambar 4.17. <i>Cubic root scaled gammatone filter bank energy</i>	40
Gambar 4.18. Hasil DCT pada <i>log mel filter bank energy</i>	40
Gambar 4.19. Hasil DCT pada <i>scaled gamma filter bank energy</i>	41
Gambar 4.20. <i>Sample Mean MFCC Feature</i>	42
Gambar 4.21. <i>Sample Mean GFCC Feature</i>	42
Gambar 4.22. Sampel kombinasi fitur MFCC+GFCC	43
Gambar 4.23. Penyusunan arsitektur CNN	45
Gambar 4.24. Kode program pelatihan model CNN.....	45
Gambar 4.25. <i>Confusion matrix CNN MFCC+GFCC batch 32</i>	51
Gambar 4.26. Grafik hasil evaluasi matriks.....	54

DAFTAR TABEL

Tabel 2.1. <i>Confusion matrix</i>	17
Tabel 2.2. Review studi literatur	18
Tabel 3.1. Jadwal penelitian.....	26
Tabel 4.1. Sampel Waveform data suara untuk setiap emosi	28
Tabel 4.2. Fitur MFCC.....	43
Tabel 4.3. Fitur GFCC	43
Tabel 4.4. Fitur MFCC+GFCC	44
Tabel 4.5. Grafik hasil proses <i>training</i> model CNN	45
Tabel 4.6. Hasil klasifikasi emosi	48
Tabel 4.7. Matriks evaluasi model klasifikasi emosi pada suara manusia.....	54