

**SKRIPSI**



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

**ENDOW BONAPEN**

**NIM. 2010511010**

**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**



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

**ENDOW BONAPEN**

**NIM. 2010511010**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA**

**2024**

ii

## PERNYATAAN ORISINALITAS

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

Nama : Endow Bonapen

NIM : 2010511010

Tanggal : 17 Januari 2024

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

Jakarta, 17 Januari 2024

Yang menyatakan,

A handwritten signature in black ink is written over a yellow rectangular meter stamp. The stamp features the Garuda Pancasila emblem at the top center, the text 'METERAL TEMPEL' below it, and the alphanumeric code 'DEFAKX644018941' at the bottom. The signature is written in a cursive style across the stamp.

(Endow Bonapen)

## PERNYATAAN PERSETUJUAN PUBLIKASI

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

Nama : Endow Bonapen

NIM : 2010511010

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 :

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

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.

Dibuat di : Depok

Pada tanggal : 17 Januari 2024

Yang menyatakan,



(Endow Bonapen)

## LEMBAR PENGESAHAN

Skripsi diajukan oleh :

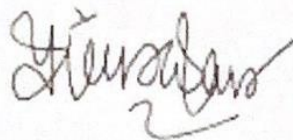
Nama : Endow Bonapen

NIM : 2010511010

Program Studi : SI Informatika

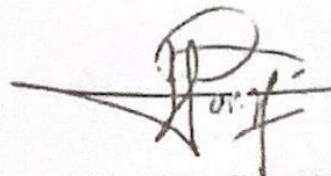
Judul Skripsi : IMPLEMENTASI METODE CONVOLUTIONAL NEURAL NETWORK  
UNTUK PENGENALAN EMOSI BERDASARKAN SUARA MANUSIA  
MENGUNAKAN REPRESENTASI CITRA SPECTROGRAM

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.



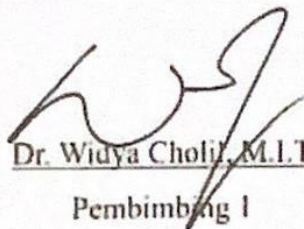
Yuni Widiastiwi, S.Kom, M.Si.

Penguji 1



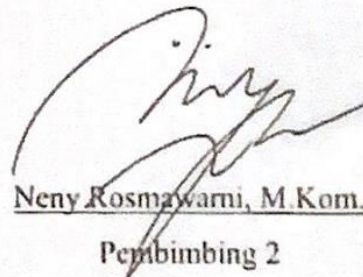
Muhammad Panji Muslim, S.Pd., M.Kom.

Penguji 2



Dr. Widya Cholil, M.I.T

Pembimbing 1



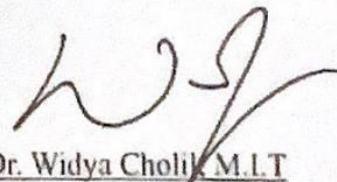
Neny Rosmawarni, M.Kom.

Pembimbing 2



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

Dekan



Dr. Widya Cholil, M.I.T

Kaprodi

Ditetapkan di : Jakarta

Tanggal Ujian : 11 Januari 2024

# 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.*

## KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Puji syukur senantiasa penulis panjatkan atas kehadiran ALLAH *سُبْحَانَهُ وَ تَعَالَى*, yang selalu melimpahkan rahmat dan karunia-NYA, sehingga Penulis dapat menyelesaikan penyusunan laporan Tugas Akhir atau Skripsi ini yang berjudul “IMPLEMENTASI METODE *CONVOLUTIONAL NEURAL NETWORK* UNTUK PENGENALAN EMOSI BERBASIS SUARA MANUSIA MENGGUNAKAN REPRESENTASI CITRA *SPECTROGRAM*”. Skripsi ini disusun sebagai syarat dalam rangka menyelesaikan studi yang sedang Penulis tempuh dalam gelar Sarjana Komputer di Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.

Dalam penyusunan Skripsi ini tidak lepas dari bimbingan dan dukungan oleh banyak pihak. Oleh karena itu penulis ingin mengucapkan terimakasih kepada:

1. Papah dan Mamah yang selalu memberikan doa dan dukungan dalam bentuk moral dan materil hingga saat ini.
2. Ibu Dr. Widya Cholil, M.I.T, selaku Dosen Pembimbing 1 Skripsi yang telah banyak memberi masukan, arahan, dan saran yang sangat berguna dan berarti bagi Penulis dalam menyelesaikan Skripsi ini.
3. Ibu Neny Rosmawarni, M.Kom, selaku Dosen Pembimbing 2 Skripsi.
4. Bapak Jayanta, S.Kom., M.Si, selaku Dosen Pembimbing Akademik.
5. Teman – teman seperjuangan saya dari tahun 2020 hingga sekarang Fiqri Fadillah, Rizky Firmansyah, dan Rizki Firmansyah yang selalu menemani dan memberikan dukungan penuh kepada penulis selama proses perkuliahan dan penyelesaian Skripsi ini.
6. Teman – teman yang baru saya jumpai di akhir masa perkuliahan Verena Nisa Silaen AKA ‘V.N.S’ (Ms. Verena atau (Daaa)), Nauval (Nompal), dan Annisa (Nis) yang telah memberikan dukungan penuh kepada Penulis dalam menyelesaikan Skripsi ini.

Bekasi, 26 Januari 2024

Endow Bonapen



## DAFTAR ISI

PERNYATAAN ORISINALITAS .....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI .....	iv
LEMBAR PENGESAHAN .....	v
ABSTRAK.....	vi
<i>ABSTRACT</i> .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL .....	xii
DAFTAR GAMBAR .....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian .....	4
1.5 Batasan Masalah.....	4
1.6 Luaran Yang Diharapkan.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Suara.....	6
2.2 Emosi .....	6
2.3 Speech Recognition.....	9
2.3.1 Pengertian Speech Recognition .....	9
2.3.2 Pengertian Speech Emotion Recognition .....	10
2.3.3 Pattern Recognition.....	12
2.3.4 <i>Mel Frequency Cepstral Coefficient (MFCC) Spectrogram</i> .....	13
2.4 Artificial Intelligence .....	18
2.5 Machine Learning .....	20
2.6 Deep Learning.....	21
2.7 Convolutional Neural Network (CNN).....	25
2.7.1 Convolutional Layer .....	27
2.7.2 Activation Function .....	29
2.7.3 Pooling Layer.....	29
2.7.4 Adam Optimizers .....	30
2.7.5 Dropout .....	32
2.7.6 Flattening .....	33

2.7.7	Fully Connected Layer.....	33
2.8	Confusion Matrix .....	33
2.9	Review Penelitian Terdahulu .....	35
BAB III METODELOGI PENELITIAN.....		38
3.1	Alur Penelitian .....	38
3.1.1	Identifikasi Masalah.....	39
3.1.2	Studi Literatur .....	39
3.1.3	Pengumpulan Data.....	39
3.1.4	Fitur Ekstraksi.....	41
3.1.5	Pembagian Data Latih, Data Validasi dan Data Uji.....	42
3.1.6	<i>Data Preparation</i> .....	43
3.1.7	Perancangan Model.....	43
3.1.8	Pelatihan Model .....	44
3.1.9	Pengujian dan Evaluasi Model .....	44
3.2	Perangkat Penelitian.....	45
3.2.1	Perangkat Keras .....	45
3.2.2	Perangkat Lunak .....	45
3.3	Jadwal Penelitian.....	45
BAB IV HASIL DAN PEMBAHASAN .....		47
4.1	Pengumpulan Data .....	47
4.1.1	<i>Sample Rate</i> .....	49
4.1.2	<i>Trimming</i> .....	50
4.1.3	<i>Reduce Noise</i> .....	50
4.2	Fitur Ekstraksi .....	51
4.3	Pembagian Data Latih, Data Validasi, dan Data Uji .....	58
4.4	Perancangan Model.....	59
4.5	Pelatihan Model .....	61
4.5.1	Pengaruh Ukuran <i>Batch Size</i> .....	71
4.5.2	Pengaruh Jumlah <i>Neuron</i> Pada <i>Hidden Layer</i> .....	72
4.5.3	Pengaruh Nilai <i>Learning Rate</i> .....	73
4.5.4	Pengaruh Jumlah <i>Epoch</i> .....	73
4.5.5	Pengaruh Pembagian Data Latih, Data Validasi, dan Data Uji.....	74
4.5.6	Pengaruh Fungsi Aktivasi .....	75
4.5.7	Pengaruh Ukuran Kernel .....	76
4.6	Pengujian dan Evaluasi Model.....	76

BAB V KESIMPULAN DAN SARAN.....	85
5.1 Kesimpulan .....	85
5.2 Saran.....	86
DAFTAR PUSTAKA.....	xiv
LAMPIRAN.....	xvii

## DAFTAR TABEL

Tabel 2.1 Parameter pada proses Convolutional Layer .....	29
Tabel 2.2 Penelitian Terdahulu .....	35
Tabel 3.1. Rincian Data Suara .....	41
Tabel 3.1 Jadwal Penelitian .....	45
Tabel 4.1 Citra MFCC Spectrogram berdasarkan kategori emosi. ....	53
Tabel 4.2 Perbandingan Rasio Data Latih, Data Validasi, dan Data Uji.....	59
Tabel 4.3 Arsitektur Model yang akan dibangun.....	60
Tabel 4.4 Hyperparameter yang digunakan dalam pelatihan dan pembangunan Model .	60
Tabel 4.5 Hasil Pelatihan Model dengan Arsitektur yang sudah ditentukan .....	61
Tabel 4.6 Penjelasan Kode Feature Learning .....	63
Tabel 4.6 Penjelasan Kode <i>Classification</i> .....	69
Tabel 4.7 <i>Hyperparameter Nonstatis</i> .....	71
Tabel 4.8 Akurasi Berdasarkan Ukuran <i>Batch Size</i> .....	71
Tabel 4.9 Akurasi Berdasarkan Jumlah <i>Neuron</i> Pada <i>Hidden Layer</i> .....	72
Tabel 4.10 Akurasi Berdasarkan Nilai <i>Learning Rate</i> .....	73
Tabel 4.11 Akurasi Berdasarkan Jumlah <i>Epoch</i> .....	74
Tabel 4.12 Akurasi Berdasarkan Rasio Pembagian <i>Dataset</i> .....	75
Tabel 4.13 Akurasi Berdasarkan Fungsi Aktivasi.....	75
Tabel 4.14 Akurasi Berdasarkan Ukuran Kernel .....	76
Tabel 4.15 Hasil <i>Confusion Matrix</i> pada Semua Label Emosi .....	77
Tabel 4.16 <i>Confusion Matrix</i> Kelas <i>Angry</i> .....	78
Tabel 4.17 <i>Confusion Matrix</i> Kelas <i>Disgust</i> .....	78
Tabel 4.18 <i>Confusion Matrix</i> Kelas <i>Fear</i> .....	79
Tabel 4.19 <i>Confusion Matrix</i> Kelas <i>Happy</i> .....	79
Tabel 4.20 <i>Confusion Matrix</i> Kelas <i>Neutral</i> .....	80
Tabel 4.21 <i>Confusion Matrix</i> Kelas <i>Sad</i> .....	80
Tabel 4.22 <i>Confusion Matrix</i> Kelas <i>Surprise</i> .....	81

## DAFTAR GAMBAR

Gambar 2.1 Proses Ekstraksi Fitur MFCC .....	14
Gambar 2.2 Mel Filter Bank .....	17
Gambar 2.3 Hierarki Ilmu Kecerdasan Buatan.....	21
Gambar 2.3 Arsitektur CNN .....	26
Gambar 2.5 Operasi <i>Max Pooling</i> .....	30
Gambar 3.1 Alur Penelitian .....	38
Gambar 3.3. Fitur Ekstraksi menggunakan MFCC <i>Spectrogram</i> .....	42
Gambar 3.4. Pembagian Dataset.....	42
Gambar 3.3. Arsitektur Klasifikasi Algoritma CNN.....	44
Gambar 4.1 Data Suara Emosi (RAVDESS) .....	48
Gambar 4.2 Data Suara Emosi (TESS).....	48
Gambar 4.3 Data Suara Emosi (SAVEE).....	49
Gambar 4.4 <i>Resampling Rate</i> data suara .....	50
Gambar 4.5 <i>Trimming</i> data sinyal suara .....	50
Gambar 4.6 <i>Reduce Noise</i> sebesar 10%.....	51
Gambar 4.7 Kode Fitur Ekstraksi kedalam representasi citra <i>spectrogram</i> menggunakan algoritma MFCC .....	51
Gambar 4.8 Kode untuk merubah <i>dataset</i> file suara menjadi <i>dataset</i> citra MFCC <i>spectrogram</i> . .....	52
Gambar 4.9 Kode untuk <i>install</i> , dan <i>import library</i> ( <i>split_folders</i> ).....	58
Gambar 4.10 Kode Tahap <i>Feature Learning</i> .....	62
Gambar 4.11 Kode Tahap <i>Classification</i> .....	68
Gambar 4.12 Proses Pemodelan Arsitektur CNN.....	70