

**PENGEMBANGAN SPEECH EMOTION RECOGNITION
BERBAHASA INDONESIA DENGAN EKSTRAKSI FITUR
HYBRID MENGGUNAKAN METODE DEEP LEARNING**

SKRIPSI

Muhammad Ilham Nurdiansyah Alfajar

20210040119



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK, KOMPUTER DAN DESAIN
UNIVERSITAS NUSA PUTRA
AGUSTUS 2025**

**PENGEMBANGAN SPEECH EMOTION RECOGNITION
BERBAHASA INDONESIA DENGAN EKSTRAKSI FITUR
HYBRID MENGGUNAKAN METODE DEEP LEARNING**

SKRIPSI

Diajukan Untuk Memenuhi Salah satu Syarat Dalam Menempuh

Gelar Sarjana Teknik Komputer

Muhammad Ilham Nurdiansyah Alfajar

20210040119



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK, KOMPUTER DAN DESAIN**

SUKABUMI

AGUSTUS 2025

PERNYATAAN PENULIS

JUDUL : PENGEMBANGAN SPEECH EMOTION RECOGNITION
BERBAHASA INDONESIA DENGAN EKSTRAKSI FITUR
HYBRID MENGGUNAKAN METODE DEEP LEARNING

NAMA : MUHAMMAD ILHAM NURDIANSYAH ALFAJAR

NIM : 20210040119

“Saya menyatakan dan bertanggungjawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai dengan bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut.”



Sukabumi, 11 Agustus 2025

Materai

Muhammad Ilham Nurdiansyah Alfajar

Penulis

PENGESAHAN SKRIPSI

JUDUL : PENGEMBANGAN SPEECH EMOTION RECOGNITION
BERBAHASA INDONESIA DENGAN EKSTRAKSI FITUR
HYBRID MENGGUNAKAN METODE DEEP LEARNING

NAMA : MUHAMMAD ILHAM NURDIANSYAH ALFAJAR

NIM : 20210040119

Skripsi ini telah diujikan dan dipertahankan di depan Dewan Penguji pada Sidang Skripsi tanggal 11 Agustus 2025, Menurut pandangan saya, Skripsi ini memadai dari segi kualitas untuk tujuan penganugerahan gelar Sarjana Komputer (S.KOM).

Sukabumi, 11 Agustus 2025

Pembimbing I

Alun Sujjada, S.Kom, M.T
NIDN. 0718108001

Ketua Penguji

Nugraha, M.Kom
NIDN. 0413098904



Pembimbing II

Ivana Lucia Kharisma, M.Kom
NIDN. 0429038002

Ketua Program Studi Teknik Informatika

Ir Somantri, S.T., M.Kom
NIDN. 0419128801

Dekan Fakultas Teknik Informatika dan Desain

Ir. Paikun S.T.M.T., IPM., ASEAN Eng
NIDN. 0402037401

HALAMAN PERSEMBAHAN

Dengan segala kerendahan hati dan rasa syukur, saya ucapkan terima kasih kepada Allah SWT yang senantiasa memberikan keberanian, ketabahan, dan semangat yang tak kenal menyerah dalam diri saya. Atas segala izin-Nya, doa-doa yang saya panjatkan selalu mendapat jawaban terbaik. Berkat rahmat dan anugerah-Nya, akhirnya saya dapat menyelesaikan tugas akhir ini dengan baik.

Dengan tulus, karya skripsi ini saya persembahkan kepada:

1. Allah SWT, segala puji bagi-Nya atas rahmat, hidayah, dan kemudahan yang selalu menyertai setiap langkah dalam kehidupan dan proses penyelesaian skripsi ini.
2. Ibuku tercinta, wanita hebat yang tidak pernah lelah mendoakan, menyemangati, dan mendampingi setiap langkah hidup saya. Terima kasih atas kasih sayang, pelukan yang menenangkan, dan nasihat-nasihat yang selalu menjadi penuntun saat saya merasa ragu dan lelah. Doa dan ketulusanmu adalah cahaya terbesar dalam hidupku.
3. Ayahku tersayang, sosok pekerja keras yang mengajarkan arti tanggung jawab, kesederhanaan, dan keteguhan. Terima kasih atas segala pengorbanan, kerja keras tanpa pamrih, serta semangat yang kau tularkan agar aku tidak pernah menyerah. Langkahmu selalu menjadi inspirasi dalam menempuh jalan ini.
4. Adikku, yang kehadirannya selalu membawa semangat dan kebahagiaan tersendiri. Terima kasih atas tawa, dukungan kecil yang bermakna, dan alasan tambahan bagi saya untuk terus berjuang.
5. Dosen Pembimbing, Bapak Alun Sujjada, S.Kom., M.T. dan Ibu Ivana Lucia Kharisma, M.Kom., atas bimbingan, arahan, dan kesabaran dalam mendampingi saya selama penyusunan skripsi ini. Ilmu dan perhatian yang diberikan sangat berarti bagi saya.
6. Seluruh dosen dan sivitas akademika Program Studi Teknik Informatika Universitas Nusa Putra, termasuk Kaprodi dan jajaran staf pengajar, yang telah memberikan ilmu dan pengalaman berharga selama masa studi saya.

7. Rekan-rekan seperjuangan, khususnya teman-teman Teknik Informatika 21F, atas kebersamaan, dukungan, dan tawa yang selalu mewarnai hari-hari selama perkuliahan. Semoga persaudaraan ini tetap terjalin erat di masa depan.
8. Teman-teman dari luar kampus, yang meskipun berada di tempat kuliah yang berbeda, tetapi tetap hadir dengan dukungan, semangat, dan inspirasi yang tak kalah berharga. Terima kasih atas hubungan pertemanan yang tulus dan saling menguatkan.
9. PT Anugrah Kreasi Digital (Mangcoding), tempat saya belajar dan bertumbuh selama masa magang. Terima kasih sebesar-besarnya kepada Bapak Nugraha, M.Kom. selaku founder atas kesempatan dan arahan yang sangat berharga, kepada seluruh staf Mangcoding atas bimbingan dan kebersamaan-nya, serta kepada teman-teman magang yang telah menjadi bagian dari perjalanan belajar dan pengalaman.
10. Seluruh pihak lainnya, yang tidak dapat saya sebutkan satu per satu, yang telah memberikan dukungan, doa, dan bantuan dalam berbagai bentuk selama masa studi dan proses penyusunan skripsi ini.

Akhir kata, saya berharap semoga karya ini dapat memberi manfaat bagi diri saya, bagi para pembaca, dan turut berkontribusi dalam perkembangan ilmu pengetahuan di masa depan. Aamiin Ya Rabbal 'Alamiin.

ABSTRACT

This research aims to develop a speech emotion recognition system in the Indonesian language using a deep learning approach with hybrid feature extraction. The extracted audio features consist of a combination of Mel-Frequency Cepstral Coefficients (MFCC), Hilbert Multispectrum, and Cochleagram, which represent the frequency, amplitude, and auditory perception aspects of a speech signal, respectively. The dataset used is IndoWaveSentiment, containing 300 speech samples categorized into five types of emotions: neutral, happy, surprised, disgusted, and disappointed. Three deep learning architectures were applied: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and a combined CNN + LSTM (Hybrid) model. Each model was trained and evaluated based on accuracy, precision, recall, F1-score, as well as Receiver Operating Characteristic (ROC) curves and Area Under Curve (AUC) values. Evaluation results showed that the hybrid model achieved the best performance with a validation accuracy of 85.00% and an AUC of 0.997, followed by the LSTM model (91.67%) and the CNN model (98.33%). As a form of real-world implementation, the emotion prediction system was developed in a local web-based interface using the Flask framework. The system allows users to upload or record audio, select the classification model, and receive emotion prediction results in real-time. Testing results indicate that all three models run effectively in the system and produce accurate and responsive predictions. This study demonstrates that the combination of hybrid feature extraction and deep learning architectures is effective in recognizing speech emotions in the Indonesian language.

Keywords: *Convolutional Neural Network, Deep Learning, Hybrid Feature Extraction, Long Short-Term Memory, Speech Emotion Recognition*

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan sistem pengenalan emosi suara (*Speech Emotion Recognition*) berbahasa Indonesia menggunakan pendekatan *deep learning* dengan ekstraksi fitur *hybrid*. Fitur suara yang digunakan terdiri dari kombinasi *Mel-Frequency Cepstral Coefficients (MFCC)*, *Multispektrum Hilbert*, dan *Cochleagram*, yang masing-masing mewakili aspek frekuensi, amplitudo, dan persepsi auditori dari sinyal suara. *Dataset* yang digunakan adalah *IndoWaveSentiment*, yang berisi 300 sampel suara dari lima jenis emosi yaitu netral, senang, terkejut, jijik, dan kecewa. Tiga arsitektur model *deep learning* diterapkan, yaitu *Convolutional Neural Network (CNN)*, *Long Short-Term Memory (LSTM)*, dan gabungan *CNN + LSTM (Hybrid)*. Masing-masing model dilatih dan dievaluasi berdasarkan metrik akurasi, *precision*, *recall*, *F1-score*, serta kurva *Receiver Operating Characteristic (ROC)* dan *Area Under Curve (AUC)*. Hasil evaluasi menunjukkan bahwa model *hybrid* memberikan performa terbaik dengan akurasi validasi sebesar 85,00% dan *AUC* sebesar 0.997, diikuti oleh model *LSTM* (91,67%) dan *CNN* (98,33%). Sebagai bentuk implementasi nyata, sistem prediksi emosi suara juga dikembangkan dalam bentuk antarmuka *web* lokal menggunakan *framework Flask*. Sistem ini memungkinkan pengguna untuk mengunggah atau merekam suara, memilih model klasifikasi, dan memperoleh hasil prediksi secara langsung. Hasil pengujian menunjukkan bahwa ketiga model dapat berjalan dengan baik dalam sistem dan memberikan hasil prediksi yang akurat serta responsif. Hasil penelitian ini membuktikan bahwa kombinasi fitur *hybrid* dan arsitektur model *deep learning*, efektif dalam mengenali emosi suara berbahasa Indonesia.

Kata Kunci: *Convolutional Neural Network, Deep Learning, Ekstraksi Fitur Hybrid, Long Short-Term Memory, Speech Emotion Recognition*

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW. Dengan penuh rasa syukur, penulis dapat menyelesaikan skripsi dengan judul “Pengembangan Speech Emotion Recognition Dengan Ekstraksi Fitur Hybrid Menggunakan Metode Deep Learning” dengan sebaik-baiknya.

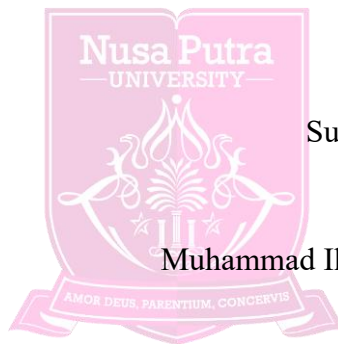
Penulis menyadari bahwa penyusunan skripsi ini masih jauh dari kesempurnaan, baik dari segi materi maupun penyajian. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang bersifat membangun guna perbaikan di masa mendatang. Semoga karya ilmiah ini dapat memberikan manfaat, khususnya bagi penulis, dan umumnya bagi para pembaca serta pihak-pihak yang berkepentingan di bidang terkait.

Dalam proses penyusunan skripsi ini, penulis mendapat banyak dukungan, bantuan, dan bimbingan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Bapak Dr. H. Kurniawan, S.T., M.H., selaku Rektor Universitas Nusa Putra;
2. Bapak Ir. H. Paikum, S.T., IPM., Asean Eng., selaku Dekan Fakultas Universitas Nusa Putra;
3. Bapak Ir. Somantri, S.T., M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Nusa Putra;
4. Bapak Alun Sujjada, S.Kom., M.T. dan Ibu Ivana Lucia Kharisma, M.Kom., selaku dosen pembimbing skripsi yang telah memberikan bimbingan, masukan, dan arahan selama proses penyusunan skripsi;
5. Bapak Alun Sujjada, S.Kom., M.T., selaku pembimbing akademik, yang selalu memberikan bimbingan, arahan, dan dukungan selama proses perkuliahan.
6. Seluruh dosen dan staf pengajar Fakultas Komputer, Teknik dan Desain Program Studi Teknik Informatika Universitas Nusa Putra atas ilmu yang telah diberikan selama masa studi;

7. Kedua orang tua, dan adik yang selalu mendukung di saat sulit untuk melangkah, menjalani hidup serta memberikan semangat dan doa.
8. keluarga, dan teman-teman yang selalu memberikan semangat, serta dukungan moral dan material kepada penulis.

Akhir kata, penulis berharap skripsi ini dapat menjadi langkah awal untuk pengembangan keilmuan yang lebih luas. Semoga Allah SWT senantiasa melimpahkan rahmat dan ridho-Nya kepada kita semua. Amin yaa Rabbal ‘Alamin.



Sukabumi, 11 Agustus 2025

Penulis
Muhammad Ilham Nurdiansyah Alfajar

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik UNIVERSITAS NUSA PUTRA, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Ilham Nurdiansyah Alfajar
NIM : 20210040119
Program Studi : Teknik Kejuruan
Jenis karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Nusa Putra Hak Bebas Royalti Eksklusif (*Non-exclusive Royalty- Free Right*) atas karya ilmiah saya yang berjudul :

PENGEMBANGAN SPEECH EMOTION RECOGNITION BERBAHASA INDONESIA DENGAN EKSTRAKSI FITUR HYBRID MENGGUNAKAN METODE DEEP LEARNING

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneklusif ini Universitas Nusa Putra berhak menyimpan, mengalih media/format-kan, 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 : SUKABUMI

Pada tanggal : 11 Agustus 2025

Yang menyatakan

Muhammad Ilham Nurdiansyah Alfajar

DAFTAR ISI

PERNYATAAN PENULIS	ii
PENGESAHAN SKRIPSI	iii
HALAMAN PERSEMBAHAN	iv
<i>ABSTRACT</i>	vi
ABSTRAK	vii
KATA PENGANTAR	viii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI	x
DAFTAR ISI	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	4
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	5
1.6. Sistematika Penelitian	6
BAB II TINJAUAN PUSTAKA	7
2.1. Penelitian Terkait	7
2.2. Landasan Teori	9
2.2.1. Metode Kuantitatif	9
2.2.2. Kecerdasan Buatan	10
2.2.3. Deep Learning	11
2.2.4. Audio	12
2.2.5. Emotion	13
2.2.6. Speech Emotion Recognition	14
2.2.7. Fast Fourier Transform	14
2.2.8. Mel Frequency Cepstral Coefficient	15
2.2.9. Multispektrum Hilbert	16
2.2.10. Cochleagram	17

2.2.11.	Convolutional Neural Network	18
2.2.12.	Long Short-Term Memory	18
2.2.13.	Dataset	19
2.2.14.	Training, Validation dan Test.....	20
2.2.15.	Confusion Matrix	21
2.2.16.	Akurasi, Precision, Recall, dan F1-Score	22
2.2.17.	AUC-ROC	24
2.2.18.	Kurva Loss dan Akurasi	25
2.3.	Kerangka Berpikir.....	27
BAB III METODE PENELITIAN		28
3.1.	Tahapan Penelitian.....	28
3.2.	Identifikasi Masalah.....	29
3.3.	Perumusan Tujuan	30
3.4.	Studi Literatur	31
3.5.	Sampel <i>Dataset</i>	32
3.5.1.	Deskripsi Dataset.....	33
3.5.2.	Visualisasi Sinyal Suara Waveplot dan Spectrogram.....	36
3.5.3.	Prosedur Perekaman	40
3.5.4.	Ukuran Dataset	41
3.5.5.	Sumber Data	41
3.5.6.	Diagram Alur Program	41
3.6.	<i>Preprocessing</i> Data.....	42
3.6.1.	Pemuatan Dataset dan Label Emosi	42
3.6.2.	Ekstraksi Fitur Hybrid	43
3.6.3.	Encoding, Normalisasi dan Split Data	45
3.7.	Pemodelan Deep Learning.....	46
3.7.1.	Arsitektur model Convolutional Neural Network	46
3.7.2.	Arsitektur model Long Short-Term Memory	47
3.7.3.	Arsitektur model Hybrid (CNN + LSTM)	48
3.8.	Pelatihan Model	50
3.9.	Evaluasi Model	50
3.10.	Pengujian Sistem.....	51

BAB IV HASIL DAN PEMBAHASAN.....	52
4.1. Klasifikasi <i>Train</i> Model.....	52
4.1.1. Train dan Validation CNN	52
4.1.2. Train dan Validation LSTM.....	52
4.1.3. Train dan Validation Hybrid (CNN + LSTM)	53
4.2. Evaluasi Arsitektur Model	53
4.2.1. Evaluasi CNN.....	54
4.2.2. Evaluasi LSTM.....	56
4.2.3. Evaluasi Hybrid (CNN + LSTM).....	58
4.3. Perbandingan Kinerja Model.....	60
4.4. Implementasi Sistem Berbasis <i>Web</i>	62
4.5. Hasil Pengujian Skenario.....	65
BAB V PENUTUP	69
5.1. Kesimpulan	69
5.2. Saran	70
DAFTAR PUSTAKA.....	71



DAFTAR GAMBAR

Gambar 2.1. Deep Learning	11
Gambar 2.2. Gambar Sinyal Audio dari Waveplot	12
Gambar 2.3. Gambar Sinyal Audio dari Spectrogram.....	12
Gambar 2.4. Speech Emotion Recognition	14
Gambar 2.5. MFCC Pipeline Steps	15
Gambar 2.6. Diagram Alur Multispektrum Hilbert.....	16
Gambar 2.7. Diagram Alur Cohleagram	17
Gambar 2.8. Algoritma CNN	18
Gambar 2.9. Algoritma LSTM	19
Gambar 2.10. Kurva Loss.....	25
Gambar 2.11. Kurva Akurasi	26
Gambar 2.12. Kerangka Berpikir	27
Gambar 3.1. Metode Penelitian	29
Gambar 3.2. bar chart Aktor.....	33
Gambar 3.3. bar chart Gender	33
Gambar 3.4. bar chart Emosi.....	34
Gambar 3.5. bar chart Intensitas.....	34
Gambar 3.6. bar chart Repetisi	35
Gambar 3.7. Struktur dan Pelabelan <i>Dataset</i>	35
Gambar 3.8. Gelombang dari Emosi Netral	36
Gambar 3.9. Spektrum dari Emosi Netral	37
Gambar 3.10. Gelombang dari Emosi Senang	37
Gambar 3.11. Spektrum dari Emosi Senang	37
Gambar 3.12. Gelombang dari Emosi Terkejut.....	38
Gambar 3.13. Spektrum dari Emosi Terkejut.....	38
Gambar 3.14. Gelombang dari Emosi Jijik	39
Gambar 3.15. Spektrum dari Emosi Jijik	39
Gambar 3.16. Gelombang dari Emosi Kecewa	39
Gambar 3.17. Spektrum dari Emosi Kecewa	40

Gambar 3.18. Prosedur Audio	40
Gambar 3.19 Diagram Alur Klasifikasi dan Sistem.....	41
Gambar 3.20. Visualisasi MFCC	43
Gambar 3.21. Visualisasi Hilbert	44
Gambar 3.22. Visualisasi Koklea	44
Gambar 4.1. <i>Curve Accuracy dan Loss - CNN</i>	52
Gambar 4.2. <i>Curve Accuracy dan Loss - LSTM</i>	53
Gambar 4.3. <i>Curve Accuracy dan Loss - Hybrid</i>	53
Gambar 4.4. <i>Confusion Matrix CNN</i>	55
Gambar 4.5. <i>Receiver Operating Characteristic CNN</i>	55
Gambar 4.6. <i>Confusion Matrix LSTM</i>	57
Gambar 4.7. <i>Receiver Operating Characteristic LSTM</i>	57
Gambar 4.8. <i>Confusion Matrix CNN+LSTM</i>	59
Gambar 4.9. <i>Receiver Operating Characteristic CNN+LSTM</i>	60
Gambar 4.10. Perbandingan Metrik Kinerja Model	61
Gambar 4.11. <i>Kurva ROC (Macro-Averaged)</i>	62
Gambar 4.12. Halaman SER	63
Gambar 4.13. Dropdown model	63
Gambar 4.14. Hasil Prediksi.....	64
Gambar 4.15. Riwayat Prediksi.....	64
Gambar 4.16. Hasil Setiap Model Untuk Rekaman	67
Gambar 4.17. Hasil Setiap Model Untuk Upload	67

DAFTAR TABEL

Tabel 2.1. Penelitian Terkait	7
Tabel 2.2. Struktur umum confusion matrix untuk dua kelas	21
Tabel 3.1. DataFrame Label <i>dataset</i>	42
Tabel 3.2. Label Emosi pada sistem.....	43
Tabel 3.3. Dimensi Fitur Hybrid	45
Tabel 3.4. Encoding Label Emosi	45
Tabel 3.5. Pembagian / Split <i>Dataset</i>	46
Tabel 3.6. Sequential CNN.....	46
Tabel 3.7. Sequential LSTM	47
Tabel 3.8. Sequential Hybrid.....	48
Tabel 4.1. <i>Metrics CNN</i>	54
Tabel 4.2. <i>Classification Report CNN</i>	54
Tabel 4.3. <i>Metrics LSTM</i>	56
Tabel 4.4. <i>Classification Report LSTM</i>	56
Tabel 4.5. <i>Metrics CNN+LSTM</i>	58
Tabel 4.6. <i>Classification Report CNN+LSTM</i>	58
Tabel 4.7. Perbandingan Kinerja Model	60
Tabel 4.8. Skenario Pengujian dengan Rekam Suara	65
Tabel 4.9. Skenario Pengujian dengan Upload Suara	66

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kemampuan mesin untuk mengenali emosi manusia melalui suara menjadi topik yang semakin penting dalam pengembangan kecerdasan buatan (*AI*). Teknologi ini telah diterapkan di berbagai sektor, seperti layanan pelanggan berbasis *AI*, deteksi kondisi kesehatan mental, hingga interaksi dalam pembelajaran daring[1]. Pada konteks layanan pelanggan, sistem yang dapat mengidentifikasi emosi pengguna mampu memberikan respon yang lebih tepat dan personal[2]. Pada bidang pendidikan, teknologi ini juga berpotensi membantu mengamati keadaan emosional siswa guna meningkatkan kualitas proses belajar mengajar[3]. Penelitian mengenai pengenalan emosi berbasis suara telah mengalami perkembangan, tantangan utama masih berkaitan dengan perbedaan intonasi, logat, serta variasi ekspresi emosi antar bahasa, termasuk bahasa Indonesia[4]. Setiap bahasa memiliki karakteristik fonetik, prosodi, dan struktur kalimat yang berbeda-beda dalam mengekspresikan emosi. Ekspresi "marah" dalam bahasa Inggris bisa disampaikan dengan intonasi tinggi dan nada cepat, sedangkan dalam bahasa Indonesia bisa saja diungkapkan dengan nada lebih datar namun dengan tekanan vokal tertentu. Keberagaman logat dan dialek dalam bahasa Indonesia, seperti logat Jawa, Sunda, atau Batak, memperkaya kompleksitas dalam mengenali emosi secara akurat dari sinyal suara. Hal ini membuat model yang dilatih pada data berbahasa asing tidak selalu dapat bekerja optimal pada data berbahasa Indonesia. Oleh karena itu, dibutuhkan penelitian khusus yang menggunakan *dataset* berbahasa Indonesia, seperti *IndoWaveSentiment*, untuk melatih dan menguji sistem pengenalan emosi agar sesuai dengan karakteristik linguistik dan emosional masyarakat Indonesia[5].

Sejumlah penelitian sebelumnya telah membuktikan bahwa emosi manusia dapat dikenali melalui sinyal suara menggunakan kombinasi metode dan fitur yang tepat. Salah satu pendekatan awal dilakukan dengan menerapkan fitur *MFCC* dan model *LSTM* untuk klasifikasi emosi dalam *dataset* berbahasa Inggris[6]. Penelitian tersebut mencapai akurasi sebesar 98,80% dan berhasil diimplementasikan ke dalam antarmuka *web* secara *real-time*, menunjukkan efektivitas *LSTM* dalam menangkap pola temporal dalam data audio[6].

Pendekatan visual berbasis spektrum juga telah dikembangkan yang dimana mengusulkan kombinasi fitur *Multispektrum Hilbert* dan *Cochleagram* sebagai representasi sinyal suara dalam bentuk citra untuk klasifikasi emosi. Dalam penelitian tersebut, *CNN* digunakan untuk memproses citra spektrum, sementara *LSTM* dimanfaatkan untuk mempelajari aspek temporal[7]. Hasilnya menunjukkan akurasi yang cukup tinggi, yaitu 90,97% untuk *CNN* dan 80,62% untuk *LSTM*, membuktikan bahwa spektrum *Hilbert* dan *Cochleagram* mampu merepresentasikan emosi secara efektif dari sinyal suara[7].

Penggunaan metode *Convolutional Neural Network (CNN)* dan *Recurrent Neural Network (RNN)* juga digunakan sebagai pembanding, serta diterapkan Teknik *Synthetic Minority Oversampling Technique (SMOTE)* untuk menangani masalah ketidakseimbangan data[8]. Pendekatan gabungan antara *Convolutional Neural Network (CNN)* dan *Decision Tree* telah menunjukkan kinerja klasifikasi yang lebih unggul dibandingkan dengan metode *deep learning* lainnya[9]. Penggunaan fitur spektral tambahan terbukti berperan penting dalam meningkatkan akurasi model *deep learning* untuk pengenalan emosi melalui suara.

Meskipun sejumlah penelitian telah dilakukan, pencapaian akurasi yang tinggi dalam pengenalan emosi berbasis suara berbahasa Indonesia masih menjadi kendala[10]. Salah satu upaya yang dilakukan untuk mengatasi hal tersebut adalah dengan mengombinasikan fitur *Multispektrum Hilbert* dan *Cochleagram* guna meningkatkan performa sistem pengenalan emosi[7]. Penerapan metode *Convolutional Neural Network (CNN)* yang dikombinasikan dengan ekstraksi fitur lainnya dapat secara signifikan meningkatkan kinerja dalam mendeteksi emosi suara dan menekankan peran penting ekstraksi fitur *Mel Frequency Cepstral Coefficient (MFCC)* dalam mengoptimalkan performa model *deep learning* untuk klasifikasi emosi[11]. Pemanfaatan *dataset IndoWaveSentiment* dalam penelitian ini dipilih karena kelengkapan dan relevansinya dalam merepresentasikan ekspresi emosional dalam bahasa Indonesia. Berbeda dengan sejumlah *dataset* sebelumnya yang umumnya berbasis bahasa Inggris, *dataset* ini menawarkan keragaman emosi dalam situasi percakapan yang lebih sesuai dengan konteks lokal. Dengan demikian, penggunaan *dataset* ini diharapkan mampu menghasilkan model yang lebih responsif terhadap nuansa fonetik dan prosodi khas bahasa Indonesia[5].

Berdasarkan identifikasi masalah dan kajian terhadap penelitian sebelumnya, studi ini bertujuan untuk merancang dan mengembangkan model pengenalan emosi berbasis suara berbahasa Indonesia dengan pendekatan *deep learning*. Model yang dibangun akan memanfaatkan teknik ekstraksi fitur *Mel Frequency Cepstral Coefficient (MFCC)*, *Multispektrum Hilbert* dan *Cochleagram* serta algoritma atau arsitektur *Convolutional Neural Network (CNN)*, *Long Short-Term Memory (LSTM)* dan gabungan *CNN-LSTM* guna meningkatkan akurasi dalam klasifikasi emosi. Pendekatan ini diharapkan dapat memberikan kontribusi yang berarti dalam kemajuan teknologi pengenalan emosi suara, terutama untuk bahasa Indonesia, serta mendukung penerapan kecerdasan buatan di berbagai sektor seperti pendidikan, *customer service*, dan pemantauan kesehatan mental.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, beberapa permasalahan utama dalam penelitian ini adalah:

1. Bagaimana membangun model pengenalan emosi berbasis suara dalam bahasa Indonesia dengan metode *deep learning* menggunakan *dataset IndoWaveSentiment* dengan 5 kelas emosi (netral, senang, terkejut, jijik, kecewa);
2. Bagaimana ekstraksi fitur *Mel Frequency Cepstral Coefficient (MFCC)*, *Multispektrum Hilbert* dan *Cochleagram* dapat meningkatkan akurasi model dalam klasifikasi emosi suara;
3. Bagaimana penggabungan ekstraksi fitur *Mel Frequency Cepstral Coefficient (MFCC)*, *Multispektrum Hilbert*, dan *Cochleagram* dapat meningkatkan akurasi model dalam mengklasifikasikan emosi suara;
4. Bagaimana perbandingan performa serta model gabungan antara *Convolutional Neural Network (CNN)* dan *Long Short-Term Memory (LSTM)* dalam mengenali emosi berdasarkan suara pada *dataset* berbahasa Indonesia;

1.3. Batasan Masalah

Untuk menjaga fokus penelitian, batasan yang diterapkan dalam penelitian ini adalah:

1. *Dataset* yang digunakan adalah *IndoWaveSentiment*, yang memiliki 5 kelas ekspresi emosional yaitu netral, senang, terkejut, jijik dan kecewa;
2. Audio hanya bisa membaca format *wav*;
3. Ekstraksi fitur yang digunakan adalah *Mel Frequency Cepstral Coefficient (MFCC)*, *Multispektrum Hilbert* dan *Cochleagram* untuk mengubah sinyal suara menjadi representasi fitur numerik;
4. Metode *deep learning* yang dibandingkan serta digabungkan adalah *Convolutional Neural Network (CNN)* dan *Long Short-Term Memory (LSTM)*;
5. Evaluasi model dilakukan berdasarkan metrik akurasi, presisi, *recall*, *F1-score*, dan *AUC-ROC* untuk mengukur performa klasifikasi emosi;
6. Penelitian ini hanya berfokus pada analisis suara tanpa mempertimbangkan data teks atau ekspresi wajah.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan model pengenalan emosi berbasis suara dalam bahasa Indonesia dengan menggunakan metode *deep learning*. Secara lebih spesifik, tujuan penelitian ini adalah:

1. Mengembangkan model pengenalan emosi suara menggunakan *dataset IndoWaveSentiment* yang mencakup 5 kelas emosi.
2. Menganalisis efektivitas ekstraksi fitur *Mel Frequency Cepstral Coefficient (MFCC)*, *Multispektrum Hilbert* dan *Cochleagram* dalam meningkatkan akurasi model *deep learning* untuk pengenalan emosi suara.
3. Menganalisis dan mengevaluasi efektivitas kombinasi ekstraksi fitur *Mel Frequency Cepstral Coefficient (MFCC)*, *Multispektrum Hilbert*, dan *Cochleagram* dalam meningkatkan akurasi model klasifikasi emosi suara.
4. Membandingkan dan mengkombinasikan performa metode *Convolutional Neural Network (CNN)* dan *Long Short-Term Memory (LSTM)* dalam mengklasifikasikan emosi berdasarkan suara.

5. Mengevaluasi tingkat akurasi model dalam membedakan kelas emosi yang terdapat dalam *dataset IndoWaveSentiment*.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat dalam berbagai aspek, baik secara akademik maupun praktis

1. Bagi Akademisi dan Peneliti

- 1) Memberikan studi perbandingan komprehensif antara metode *Convolutional Neural Network (CNN)* dan *Long Short-Term Memory (LSTM)*, dalam pengenalan emosi berbasis suara, khususnya dengan data berbahasa Indonesia.
- 2) Menjadi referensi dalam pengembangan penelitian bidang *speech emotion recognition* dengan efektivitas penggunaan fitur *Mel Frequency Cepstral Coefficient (MFCC)*, *Multispektrum Hilbert*, dan *Cochleagram* sebagai teknik ekstraksi fitur.
- 3) Membuka peluang penelitian lanjutan yang dapat mengeksplorasi integrasi teknik *deep learning* lainnya atau pengembangan *dataset* lokal yang lebih representatif.

2. Bagi Praktisi dan Pengembang Sistem AI

- 1) Mendukung pengembangan sistem interaksi berbasis suara yang lebih cerdas dan responsif secara emosional, seperti asisten *virtual* atau sistem pembelajaran daring yang mampu memahami dan menyesuaikan respons berdasarkan kondisi emosional pengguna.
- 2) Menyediakan model analisis sentimen berbasis suara yang lebih akurat dan dapat diterapkan dalam berbagai bidang praktis, seperti layanan pelanggan dan psikologi digital.

3. Bagi Masyarakat Umum

- 1) Memberikan solusi dalam berbagai sektor, seperti pendidikan daring dan layanan konseling *virtual*, yang membutuhkan kemampuan sistem untuk mengenali dan merespons emosi secara tepat.

1.6. Sistematika Penelitian

Adapun sistematika penulisan dalam penelitian ini adalah sebagai berikut

BAB I PENDAHULUAN

Berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

BAB II TINJAUAN PUSTAKA

Menyajikan teori-teori yang mendasari penelitian ini, termasuk teori pengenalan emosi berbasis suara, *machine learning*, dan teknik-teknik yang digunakan.

BAB III METODE PENELITIAN

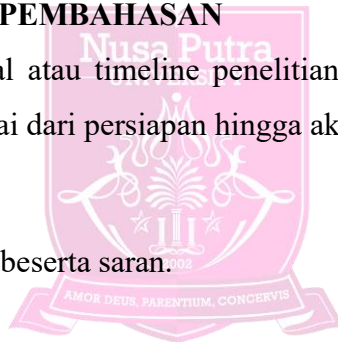
Menguraikan metode yang digunakan dalam penelitian ini, mulai dari pengumpulan data, pemilihan model, hingga evaluasi hasil.

BAB IV HASIL DAN PEMBAHASAN

Memperlihatkan jadwal atau timeline penelitian yang mencakup seluruh tahapan penelitian mulai dari persiapan hingga akhir.

BAB V PENUTUP

Kesimpulan penelitian beserta saran.



BAB V PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian dan implementasi sistem pengenalan emosi suara berbahasa Indonesia, diperoleh beberapa kesimpulan sebagai berikut :

1. Penelitian ini berhasil membangun sistem pengenalan emosi suara berbahasa Indonesia menggunakan metode *deep learning*, dengan menggabungkan tiga jenis fitur audio yaitu *MFCC*, *Multispektrum Hilbert*, dan *Cochleagram*, dalam satu representasi fitur *hybrid*.
2. Tiga model arsitektur *deep learning* digunakan, yaitu *CNN*, *LSTM*, dan *Hybrid CNN + LSTM*, dengan tujuan mengevaluasi dan membandingkan performa masing-masing dalam mengklasifikasikan emosi suara.
3. Berdasarkan hasil pelatihan dan evaluasi, *Model CNN* menunjukkan performa terbaik dengan akurasi validasi 98,33% dan *AUC-ROC* 1.00. *Model LSTM* memberikan hasil yang baik dengan akurasi validasi 91,67% dan *AUC-ROC* 1.00, menandakan kemampuannya dalam menangkap informasi temporal dari sinyal suara. *Model Hybrid* meraih akurasi validasi 85,00% dan *AUC-ROC* 0.979, namun performanya masih di bawah dua model lainnya, ini diakibatkan oleh arsitektur yang lebih kompleks dan keterbatasan jumlah data.
4. Ketiga model berhasil diintegrasikan ke dalam sistem *web* lokal, yang memungkinkan pengguna untuk memilih model, merekam atau mengunggah suara, dan langsung memperoleh hasil prediksi emosi secara responsif.

Secara keseluruhan, pendekatan ekstraksi fitur *hybrid* yang dipadukan dengan model *CNN* dan *LSTM* terbukti efektif dan efisien dalam mengenali emosi suara berbahasa Indonesia.

5.2. Saran

Berdasarkan pelaksanaan dan hasil penelitian ini, terdapat beberapa saran yang dapat menjadi pertimbangan untuk pengembangan penelitian selanjutnya :

1. Perluasan dan variasi *dataset*, *Dataset IndoWaveSentiment* yang digunakan masih terbatas dari segi jumlah dan variasi. Disarankan untuk menggunakan *dataset* yang lebih besar dan beragam seperti aksan dan kondisi lingkungan agar model dapat melakukan generalisasi lebih baik terhadap data nyata.
2. Pengembangan arsitektur model, penelitian selanjutnya dapat mengeksplorasi variasi arsitektur dalam lingkup *CNN* dan *LSTM*, seperti penambahan lapisan, penyesuaian parameter, penggunaan *BiLSTM*, atau penerapan *Attention* sederhana guna meningkatkan akurasi klasifikasi emosi.
3. Implementasi ke sistem *real-time*, sistem yang masih berbasis lokal dapat dikembangkan menjadi aplikasi *real-time* dan terintegrasi ke *platform* seperti *chatbot* atau layanan interaktif agar mampu merespons emosi pengguna secara langsung.
4. Pengujian di kondisi nyata, diperlukan pengujian tambahan dalam kondisi dunia nyata yang mengandung *noise* atau interaksi alami, untuk mengukur ketahanan dan keandalan model secara lebih menyeluruh.

DAFTAR PUSTAKA

- [1] E. Saravia, H.-C. T. Liu, Y.-H. Huang, J. Wu, and Y.-S. Chen, "CARER: Contextualized Affect Representations for Emotion Recognition."
- [2] A. Nugroho, D. Pramono Adi, and A. Bimo Gumelar, "Chatbot Untuk Customer Service Berbasis Teks dan Suara pada Sistem Manajemen Pemesanan (OMS) Menggunakan Platform Android," *REPOSITOR*, vol. 2, no. 6, pp. 683–690, Jun. 2020, doi: <https://doi.org/10.22219/repositor.v2i6.30745>.
- [3] P. Pratama Hendri, T. Wibowo, T. Tan, and Herman, "Emotion Recognition untuk mendukung Kegiatan Pembelajaran yang Menyenangkan: Studi Literatur," Jul. 2024. doi: <https://doi.org/10.58520/jddat.v3i2.63>.
- [4] M. Arista Jiwanggi and M. Adriani, "Speech-Emotion Detection in an Indonesian Movie," 2020.
- [5] A. Bustamin, A. M. Rizky, E. Warni, I. Sari Areni, and Indrabayu, "IndoWaveSentiment: Indonesian audio dataset for emotion classification," Jun. 2024, doi: 10.17632/j9ytfdz27/1.
- [6] Ivana Lucia Kharisma, Somantri, Dila Aura Futri, Muhammad Ilham Nurdiansyah Alfajar, Tirawati, and Ika, "Klasifikasi Jenis Emosi Melalui Pengenalan Suara Menggunakan Model LSTM dan Ekstraksi Fitur MFCC," <https://ieeexplore.ieee.org/document/10956929>, Apr. 2025, doi: 10.1109/ICIC64337.2024.10956929.
- [7] A. Bimo *et al.*, "Kombinasi Fitur Multispektrum Hilbert dan Cochleagram untuk Identifikasi Emosi Wicara (Spectrum Features Combination of Hilbert and Cochleagram for Speech Emotions Identification)," May 2020.
- [8] F. KASYIDI, R. ILYAS, and N. M. ANNISA, "Peningkatan Kemampuan Pengenalan Emosi Melalui Suara dalam Bahasa Indonesia," *MIND Journal*, vol. 6, no. 2, pp. 194–204, Dec. 2021, doi: 10.26760/mindjournal.v6i2.194-204.
- [9] N. Damodar, H. Y. Vani, and M. A. Anusuya, "Voice emotion recognition using CNN and decision tree," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 4245–4249, doi: 10.35940/ijitee.L2698.1081219.

- [10] M. Novela and T. Basaruddin, "DATASET SUARA DAN TEKS BERBAHASA INDONESIA PADA REKAMAN PODCAST DAN TALK SHOW," *Agustus*, vol. 11, no. 2, pp. 61–66, Aug. 2021.
- [11] Y. K. Aini, T. B. Santoso, and T. Dutono, "Pemodelan CNN Untuk Deteksi Emosi Berbasis Speech Bahasa Indonesia," 2021. [Online]. Available: <https://jurnal.pcr.ac.id/index.php/jkt/>
- [12] S. Madanian *et al.*, "Speech emotion recognition using machine learning — A systematic review," Nov. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.iswa.2023.200266.
- [13] A. Khan *et al.*, "Automatic Speech Emotion Recognition using Mel Frequency Cepstrum Co-efficient and Machine Learning Technique," Mar. 2021. [Online]. Available: <https://www.researchgate.net/publication/350467478>
- [14] Y. Yurtay, H. Demirci, H. Tiryaki, and T. Altun, "Emotion Recognition on Call Center Voice Data," *Applied Sciences (Switzerland)*, vol. 14, no. 20, Oct. 2024, doi: 10.3390/app14209458.
- [15] R. Islam, E. Abdel-Raheem, and M. Tarique, "Cochleagram to Recognize Dysphonia: Auditory Perceptual Analysis for Health Informatics," *IEEE Access*, vol. 12, pp. 59198–59210, 2024, doi: 10.1109/ACCESS.2024.3392808.
- [16] Ainurrochman, D. Pramono Adi, and A. Bimo Gumelar, "Terakreditasi SINTA Peringkat 2 Deteksi Emosi Wicara pada Media On-Demand menggunakan SVM dan LSTM," *masa berlaku mulai*, vol. 1, no. 3, pp. 799–804, 2020.
- [17] Y. B. Widodo, S. Sibuea, and M. Narji, "Kecerdasan Buatan dalam Pendidikan: Meningkatkan Pembelajaran Personalisasi," *Jurnal Teknologi Informatika dan Komputer*, vol. 10, no. 2, pp. 602–615, Oct. 2024, doi: 10.37012/jtik.v10i2.2324.
- [18] J. Thangaiyan, "EMOTION SPEECH RECOGNITION USING MFCC AND RESIDUAL PHASE IN ARTIFICIAL NEURAL NETWORK." [Online]. Available: <http://www.ijerst.com/currentissue.php>
- [19] A. W. Tusa Bagus, "KLASIFIKASI EMOSI PADA TEKS MENGGUNAKAN METODE DEEP LEARNING," 2022.
- [20] A. Raup, W. Ridwan, Y. Khoeriyah, Q. Yuliati Zaqiah, and U. Islam Negeri Sunan Gunung Djati Bandung, "Deep Learning dan Penerapannya dalam

- Pembelajaran,” Sep. 2022. [Online]. Available: <http://Jiip.stkipyapisdampu.ac.id>
- [21] J. Nurhakiki *et al.*, “Studi Kepustakaan: Pengenalan 4 Algoritma Pada Pembelajaran Deep Learning Beserta Implikasinya,” *Jurnal Pendidikan Berkarakter*, no. 1, pp. 270–281, 2024, doi: 10.51903/pendekar.v2i1.598.
 - [22] S. Das, S. Pal, and M. Mitra, “Supervised model for Cochleagram feature based fundamental heart sound identification,” *Biomed Signal Process Control*, vol. 52, pp. 32–40, doi: 10.1016/j.bspc.2019.01.028.
 - [23] C. Lan, Y. Q. Wang, L. Zhang, C. Liu, and X. Lin, “Research on Speech Enhancement Algorithm of Multiresolution Cochleagram Based on Skip Connection Deep Neural Network,” *J Sens*, vol. 2022, 2022, doi: 10.1155/2022/5208372.
 - [24] R. Fazira, D. Yudistira, and L. Sofinah Harahap, “Evaluasi Kinerja Model RNN & LSTM untuk Prediksi Magnitude Gempa di Indonesia,” *Mars : Jurnal Teknik Mesin, Industri, Elektro Dan Ilmu Komputer*, vol. 2, no. 6, pp. 62–75, Nov. 2024, doi: 10.61132/mars.v2i6.498.
 - [25] Y. Liu, A. Chen, G. Zhou, J. Yi, J. Xiang, and Y. Wang, “Combined CNN LSTM with attention for speech emotion recognition based on feature-level fusion,” *Multimed Tools Appl*, vol. 83, no. 21, pp. 59839–59859, Jan. 2024, doi: 10.1007/s11042-023-17829-x.
 - [26] F. Kasyidi and D. P. Lestari, “Identification of four class emotion from Indonesian spoken language using acoustic and lexical features,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing. doi: 10.1088/1742-6596/971/1/012048.
 - [27] M. Lech, M. Stolar, C. Best, and R. Bolia, “Real-Time Speech Emotion Recognition Using a Pre-trained Image Classification Network: Effects of Bandwidth Reduction and Companding,” *Front Comput Sci*, vol. 2, May 2020, doi: 10.3389/fcomp.2020.00014.
 - [28] B. Vimal, M. Surya, Darshan, V. S. Sridhar, and A. Ashok, “MFCC Based Audio Classification Using Machine Learning,” in *2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/ICCCNT51525.2021.9579881.
 - [29] P. P. Singh and P. Rani, “An Approach to Extract Feature using MFCC.” [Online]. Available: www.iosrjen.org

- [30] I. Dewa Agung Adwitya Prawangsa and A. Eka Karyawati, "Penerapan Metode MFCC dan LSTM untuk Speech Emotion Recognition," vol. 12, no. 4, pp. 2654–5101, May 2024, [Online]. Available: www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tess.
- [31] J. Luque, D. Anguita, F. Pérez, and R. Denda, "Spectral analysis of electricity demand using hilbert–huang transform," *Sensors (Switzerland)*, vol. 20, no. 10, May 2020, doi: 10.3390/s20102912.
- [32] S. Abedi and M. S. Moslehian, "Extensions of the Hilbert-multi-norm in Hilbert C^* -modules," May 2024, doi: 10.1007/s11117-022-00960-8.
- [33] H. Mukhtar, Y. Rizki, F. Apri Wenando, and M. Abdul Al Aziz, "Prediksi Kunjungan Wisatawan dengan Reduksi Noise pada Google Trends menggunakan Hilbert-Huang Transform dan Long Short-Term Memory", [Online]. Available: <https://www.bps.go.id/indicator/16/74/2/jumlah->
- [34] M. J. Afroni and O. Melfazen, "Deteksi dan Klasifikasi Sinyal Listrik Non-Stasioner yang Mengandung Masalah Kualitas Daya Menggunakan Transformasi Hilbert Huang," *Jurnal Elektro*, vol. 2, no. 2, p. 7, doi: 10.30736/je.v2i2.81.
- [35] F. Haider and S. Luz, "Affect Recognition through Scalogram and Multi-resolution Cochleagram Features," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, International Speech Communication Association, 2021, pp. 581–585. doi: 10.21437/Interspeech.2021-1761.
- [36] Z. Peng, W. He, Y. Li, Y. Du, and J. Dang, "Multi-Level Attention-Based Categorical Emotion Recognition Using Modulation-Filtered Cochleagram," *Applied Sciences*, vol. 13, no. 11, p. 6749, Jun. 2023, doi: 10.3390/app13116749.
- [37] M. M. Taye, "Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions," Mar. 01, 2023, *MDPI*. doi: 10.3390/computation11030052.
- [38] A. Al Bataineh, D. Kaur, M. Al-khassaweneh, and E. Al-sharoha, "Automated CNN Architectural Design: A Simple and Efficient Methodology for Computer Vision Tasks," *Mathematics*, vol. 11, no. 5, Mar. 2023, doi: 10.3390/math11051141.
- [39] S. Cong and Y. Zhou, "A review of convolutional neural network architectures and their optimizations," *Artif Intell Rev*, vol. 56, no. 3, pp. 1905–1969, Mar. 2023, doi: 10.1007/s10462-022-10213-5.

- [40] D. M. Gunarto, S. Sa'adah, and D. Q. Utama, "Predicting Cryptocurrency Price Using RNN and LSTM Method," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 1, pp. 1–8, Mar. 2023, doi: 10.32736/sisfokom.v12i1.1554.
- [41] S. S. Nurashila, F. Hamami, and T. F. Kusumasari, "PERBANDINGAN KINERJA ALGORITMA RECURRENT NEURAL NETWORK (RNN) DAN LONG SHORT-TERM MEMORY (LSTM): STUDI KASUS PREDIKSI KEMACETAN LALU LINTAS JARINGAN PT XYZ," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 3, pp. 864–877, Aug. 2023, doi: 10.29100/jupi.v8i3.3961.
- [42] N. Juan Calvin Pradiptamurty, H. Khayyiroh Shafro, M. Sihabudin Al Qurtubi, G. Alberta Tambahjong, and M. Maryamah, "Speech Emotion Recognition (SER) dengan Metode Bidirectional LSTM," *Seminar Nasional Sains Data*, vol. 2023, 2023.
- [43] F. Jonatan Tanudjaja, E. Y. Puspaningrum, and V. Via, "Klasifikasi Jenis Emosi Melalui Ucapan Menggunakan Metode Convolutional Neural Network Type Of Emotions Classification Based On Speech Using Convolutional Neural Network Method," *Online) Teknologi: Jurnal Ilmiah Sistem Informasi*, vol. 13, no. 2, pp. 1–11, Jul. 2023, doi: 10.26594/teknologi.v13i1.3740.
- [44] Hami Ismail, "Project: Speech Emotion Recognition using CNN with CREMA-D dataset using Google Colab," <https://hami-asmai.medium.com/project-speech-emotion-recognition-using-cnn-with-crema-d-dataset-using-google-colab-8cb4fdbd5044>.
- [45] Rendi Nurcahyo and Mohammad Iqbal, "Pengenalan Emosi Pembicara Menggunakan Convolutional Neural Networks," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 1, pp. 115–122, Feb. 2022, doi: 10.29207/resti.v6i1.3726.
- [46] Ajrana, A. Akbar, and A. Lawi, "Konferensi Nasional Ilmu Komputer (KONIK) 2021 Implementasi Algoritma Deep Artificial Neural Network Menggunakan Mel Frequency Cepstrum Coefficient Untuk Klasifikasi Audio Emosi Manusia," 2021, [Online]. Available: <https://www.kaggle.com/ejlok1/toronto->
- [47] D. Arfian Lazzuardhy and B. Henryranu Prasetyo, "Sistem Pengenalan Intensitas Emosi Sedih melalui Ucapan menggunakan Ekstraksi Bark-Frequency Cepstral Coefficient dan K-Nearest Neighbor berbasis Raspberry Pi 4," Aug. 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>

- [48] R. Al Kiramy, I. Permana, A. Marsal, M. R. Munzir, and M. Megawati, “Perbandingan Performa Algoritma RNN dan LSTM dalam Prediksi Jumlah Jamaah Umrah pada PT. Hajar Aswad,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1224–1234, Jul. 2024, doi: 10.57152/malcom.v4i4.1373.
- [49] T. B. Sianturi, I. Cholissodin, and N. Yudistira, “Penerapan Algoritma Long Short-Term Memory (LSTM) berbasis Multi Fungsi Aktivasi Terbobot dalam Prediksi Harga Ethereum,” 2023. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [50] W. Afandi, S. N. Saputro, A. M. Kusumaningrum, H. Ardiansyah, M. H. Kafabi, and S. Sudianto, “Klasifikasi Judul Berita Clickbait menggunakan RNN-LSTM,” vol. 7, no. 2, 2022.
- [51] C. R. Wairata, E. R. Swedia, and M. Cahyanti, “PENGKLASIFIKASIAN GENRE MUSIK INDONESIA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK,” *Sebatik*, vol. 25, no. 1, pp. 255–261, Jun. 2021, doi: 10.46984/sebatik.v25i1.1286.
- [52] N. Selle, N. Yudistira, and C. Dewi, “PERBANDINGAN PREDIKSI PENGGUNAAN LISTRIK DENGAN MENGGUNAKAN METODE LONG SHORT TERM MEMORY (LSTM) DAN RECURRENT NEURAL NETWORK (RNN),” vol. 9, no. 1, pp. 155–162, 2022, doi: 10.25126/jtiik.202295585.

